Crystalizing the Genetic Code

L. Frappat[a], A. Sciarrino[b,a], P. Sorba[a]

[a] *Laboratoire d'Annecy-le-Vieux de Physique Théorique LAPTH*
CNRS, UMR 5108, associée à l'Université de Savoie
BP 110, F-74941 Annecy-le-Vieux Cedex, France

[b] *Permanent adress: Dipartimento di Scienze Fisiche, Università di Napoli "Federico II"*
and I.N.F.N., Sezione di Napoli
Complesso di Monte S. Angelo, Via Cintia, I-80126 Napoli, Italy

Abstract

New developments are presented in the framework of the model introduced by the authors in refs. [1, 2] and in which nucleotides as well as codons are classified in crystal bases of the quantum group $U_q(sl(2) \oplus sl(2))$ in the limit $q \to 0$. An operator which gives the correspondence between the amino-acids and the codons is now obtained for any known genetic code. The free energy released by base pairing of dinucleotides as well as the relative hydrophilicity and hydrophobicity of the dinucleosides are also computed. For the vertebrate series, a universal behaviour in the ratios of codon usage frequencies is put in evidence and is shown to fit nicely in our model. Then a first attempt to represent the mutations relative to the deletion of a pyrimidine by action of a suitable crystal spinor operator is proposed. Finally recent theoretical descriptions are reviewed and compared with our model.

PACS number: 87.10.+e, 02.10.-v

LAPTH-787/00
DSF-TH-9/00
physics/0003037
March 2000

1 Introduction

Among the numerous and important questions offered to the physicist by the sciences of life, the ones relative to the genetic code present a particular interest. Indeed, in addition to the fundamental importance of this domain, the DNA structure on the one hand and the mechanism of polypeptid fixation from codons on the other hand possess appealing aspects for the theorist. Let us, in a brief summary, select some essential features [3]. First, as well known, the DNA macromolecule is constituted by two linear chains of nucleotides in a double helix shape. There are four different nucleotides, characterized by their bases: adenine (A) and guanine (G) deriving from purine, and cytosine (C) and thymine (T) coming from pyrimidine. Note also that an A (resp. T) base in one strand is connected with two hydrogen bonds to a T (resp. A) base in the other strand, while a C (resp. G) base is related to a G (resp. C) base with three hydrogen bonds. The genetic information is transmitted to the cytoplasm via the messenger ribonucleic acid or mRNA. During this operation, called transcription, the A, G, C, T bases in the DNA are associated respectively to the U, C, G, A bases, U denoting the uracile base. Then it will be through a ribosome that a triplet of nucleotides or codon will be related to an amino-acid. More precisely, a codon is defined as an ordered sequence of three nucleotides, e.g. AAG, ACG, etc., and one enumerates in this way $4 \times 4 \times 4 = 64$ different codons. Following the universal eukariotic code (see Table 4), 61 of such triplets can be connected in an unambiguous way to the amino-acids, except the three following triplets UAA, UAG and UGA, which are called non-sense or stop-codons, the role of which is to stop the biosynthesis. Indeed, the genetic code is the association between codons and amino-acids. But since one distinguishes only 20 amino-acids[1] related to the 61 codons, it follows that the genetic code is degenerated. Still considering the standard eukariotic code, one observes sextets, quadruplets, triplets, doublets and singlets of codons, each multiplet corresponding to a specific amino-acid. Such a picture naturally suggests to look for an underlying symmetry able to describe the observed structure in multiplets, in the spirit of dynamical symmetry scheme which has proven so powerful in atomic, molecular and nuclear physics. We review at the end of this paper these recent approaches.

In refs. [1, 2] we have proposed a mathematical framework in which the codons appear as composite states of nucleotides. The four nucleotides being assigned to the fundamental irreducible representation of the quantum group $\mathcal{U}_q(sl(2) \oplus sl(2))$ in the limit $q \to 0$, the codons are obtained as tensor product of nucleotides. Indeed, the properties of quantum group representations in the limit $q \to 0$, or crystal basis, are well adapted to take into account the nucleotide ordering. Then properties of this model have been considered. We will generalize some of them in the following and also propose new developments.

[1] Alanine (Ala), Arginine (Arg), Asparagine (Asn), Aspartic acid (Asp), Cysteine (Cys), Glutamine (Gln), Glutamic acid (Glu), Glycine (Gly), Histidine (His), Isoleucine (Ile), Leucine (Leu), Lysine (Lys), Methionine (Met), Phenylalanine (Phe), Proline (Pro), Serine (Ser), Threonine (Thr), Tryptophane (Trp), Tyrosine (Tyr), Valine (Val).

The paper is organized as follows. We start in sect. 2 by recalling the main aspects of the model. In sect. 3 we build out of the generators of $\mathcal{U}_{q\to 0}(sl(2)\oplus sl(2))$ a reading operator, which gives the correct correspondence between codons and amino-acids for each of the 12 presently known genetic codes. This construction generalizes in a synthetical way the one started in [1] for the eukariotic and vertebrate mitochondrial codes, the different reading operators acting on codons and providing the same eigenvalue for a given amino-acid whatever the considered code. In sect. 4 some physical properties of dinucleotide states are fitted. In sect. 5, we analyze ratios of codon usage frequency for several biological species belonging to the vertebrate class and put in evidence a universal behaviour, which fits naturally in our model. In sect. 6, making use of the general crystal basis mathematical framework, we represent the mutation induced by the deletion of a pyrimidine by the action of a suitable crystal spinor operator. In sect. 7 we review and compare with our model the recent symmetry approaches to the genetic code. Finally in sect. 8 we give a few conclusions and discuss some directions of future developments.

2 The Model

We consider the four nucleotides as basic states of the $(\frac{1}{2},\frac{1}{2})$ representation of the $\mathcal{U}_q(sl(2)\oplus sl(2))$ quantum enveloping algebra in the limit $q\to 0$. A triplet of nucleotides will then be obtained by constructing the tensor product of three such four-dimensional representations. Actually, this approach mimicks the group theoretical classification of baryons made out from three quarks in elementary particles physics, the building blocks being here the A, C, G, T/U nucleotides. The main and essential difference stands in the property of a codon to be an *ordered* set of three nucleotides, which is not the case for a baryon.

Constructing such pure states is made possible in the framework of any algebra $\mathcal{U}_{q\to 0}(\mathcal{G})$ with $\mathcal{G}$ being any (semi)-simple classical Lie algebra owing to the existence of a special basis, called crystal basis, in any (finite dimensional) representation of $\mathcal{G}$. The algebra $\mathcal{G} = sl(2)\oplus sl(2)$ appears the most natural for our purpose. The complementary rule in the DNA–mRNA transcription may suggest to assign a *quantum number* with opposite values to the couples (A,T/U) and (C,G). The distinction between the purine bases (A,G) and the pyrimidine ones (C,T/U) can be algebraically represented in an analogous way. Thus considering the fundamental representation $(\frac{1}{2},\frac{1}{2})$ of $sl(2)\oplus sl(2)$ and denoting $\pm$ the basis vector corresponding to the eigenvalues $\pm\frac{1}{2}$ of the J_3 generator in any of the two $sl(2)$ corresponding algebras, we will assume the following "biological" spin structure:

$$\begin{tikzcd}[column sep=huge, row sep=large] C \equiv (+,+) \arrow[r, leftrightarrow, "sl(2)_H"] \arrow[d, leftrightarrow, "sl(2)_V"'] & U \equiv (-,+) \arrow[d, leftrightarrow, "sl(2)_V"] \\ G \equiv (+,-) \arrow[r, leftrightarrow, "sl(2)_H"'] & A \equiv (-,-) \end{tikzcd} \tag{1}$$

(2)

the subscripts H (:= horizontal) and V (:= vertical) being just added to specify the algebra.

Now, we consider the representations of $\mathcal{U}_q(sl(2))$ and more specifically the crystal bases obtained when $q \to 0$. Introducing in $\mathcal{U}_{q\to 0}(sl(2))$ the operators J_+ and J_- after modification of the corresponding simple root vectors of $\mathcal{U}_q(sl(2))$, a particular kind of basis in a $\mathcal{U}_q(sl(2))$-module can be defined. Such a basis is called a crystal basis and carries the property to undergo in a specially simple way the action of the J_+ and J_- operators: as an example, for any couple of vectors u, v in the crystal basis $\mathcal{B}$, one gets $u = J_+v$ if and only if $v = J_-u$. More interesting for our purpose is the crystal basis in the tensorial product of two representations. Then the following theorem holds [4] (written here in the case of $sl(2)$):

Theorem 1 (Kashiwara) *Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be the crystal bases of the M_1 and M_2 $\mathcal{U}_{q\to 0}(sl(2))$-modules respectively. Then for $u \in \mathcal{B}_1$ and $v \in \mathcal{B}_2$, we have:*

$$J_-(u \otimes v) = \left\{ \begin{array}{ll} J_-u \otimes v & \exists\, n \geq 1 \text{ such that } J_-^n u \neq 0 \text{ and } J_+v = 0 \\ u \otimes J_-v & \text{otherwise} \end{array} \right. \tag{3}$$

$$J_+(u \otimes v) = \left\{ \begin{array}{ll} u \otimes J_+v & \exists\, n \geq 1 \text{ such that } J_+^n v \neq 0 \text{ and } J_-u = 0 \\ J_+u \otimes v & \text{otherwise} \end{array} \right. \tag{4}$$

Note that the tensor product of two representations in the crystal basis is not commutative. However, in the case of our model, we only need to construct the n-fold tensor product of the fundamental representation $(\frac{1}{2}, \frac{1}{2})$ of $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$ by itself, thus preserving commutativity and associativity.

Let us insist on the choice of the crystal basis, which exists only in the limit $q \to 0$. In a codon the order of the nucleotides is of fundamental importance (e.g. CCU → `Pro`, CUC → `Leu`, UCC → `Ser`). If we want to consider the codons as composite states of the (elementary) nucleotides, this surely cannot be done in the framework of Lie (super)algebras. Indeed in the Lie theory, the composite states are obtained by performing tensor products of the fundamental irreducible representations. They appear as linear combinations of the elementary states, with symmetry properties determined from the tensor product (i.e. for $sl(n)$, by the structure of the corresponding Young tableaux). On the contrary the crystal basis provides us with the mathematical structure to build composite states as *pure* states, characterized by the order of the constituents. In order to dispose of such a basis, we need to consider the limit $q \to 0$. Note that in this limit we do not deal anymore either with a Lie algebra or with an universal deformed enveloping algebra.

To represent a codon, we have to perform the tensor product of three $(\frac{1}{2}, \frac{1}{2})$ representations of $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$. However, it is well-known (see Tables 4) that in a multiplet of codons relative to a specific amino-acid, the two first bases constituent of a codon are "relatively stable", the degeneracy being mainly generated by the third nucleotide. We consider first the tensor product:

$$(\tfrac{1}{2}, \tfrac{1}{2}) \otimes (\tfrac{1}{2}, \tfrac{1}{2}) = (1,1) \oplus (1,0) \oplus (0,1) \oplus (0,0) \tag{5}$$

where inside the parenthesis, $j = 0, \frac{1}{2}, 1$ is put in place of the $2j+1 = 1, 2, 3$ respectively dimensional $sl(2)$ representation. We get, using Theorem 1, the following tableau:

$$\begin{array}{lcccc}
\rightarrow\ su(2)_H & (0,0) & (\mathrm{CA}) & (1,0) & (\ \mathrm{CG}\ \ \mathrm{UG}\ \ \mathrm{UA}\ ) \\
\downarrow & & & & \\
su(2)_V & (0,1) & \begin{pmatrix} \mathrm{CU} \\ \mathrm{GU} \\ \mathrm{GA} \end{pmatrix} & (1,1) & \begin{pmatrix} \mathrm{CC} & \mathrm{UC} & \mathrm{UU} \\ \mathrm{GC} & \mathrm{AC} & \mathrm{AU} \\ \mathrm{GG} & \mathrm{AG} & \mathrm{AA} \end{pmatrix}
\end{array}$$

From Table 4, the dinucleotide states formed by the first two nucleotides in a codon can be put in correspondence with quadruplets, doublets or singlets of codons relative to an amino-acid. Note that the sextets (resp. triplets) are viewed as the sum of a quadruplet and a doublet (resp. a doublet and a singlet). Let us define the "charge" Q of a dinucleotide state by

$$Q = J_{H,3}^{(1)} + J_{H,3}^{(2)} + J_{V,3}^{(2)} \tag{6}$$

where the superscript (1) or (2) denotes the position of a codon in the dinucleotide state.
The dinucleotide states are then split into two octets with respect to the charge Q: the eight *strong* dinucleotides associated to the quadruplets (as well as those included in the sextets) of codons satisfy $Q > 0$, while the eight *weak* dinucleotides associated to the doublets (as well as those included in the triplets) and eventually to the singlets of codons satisfy $Q < 0$. Let us remark that by the change $C \leftrightarrow A$ and $U \leftrightarrow G$, which is equivalent to the change of the sign of $J_{3,\alpha}$ or to reflexion with respect to the diagonals of the eq.(2), the 8 strong dinucleotides are transformed into weak ones and vice-versa.

If we consider the three-fold tensor product, the content into irreducible representations of $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$ is given by:

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2}) = (\tfrac{3}{2},\tfrac{3}{2}) \oplus 2\,(\tfrac{3}{2},\tfrac{1}{2}) \oplus 2\,(\tfrac{1}{2},\tfrac{3}{2}) \oplus 4\,(\tfrac{1}{2},\tfrac{1}{2}) \tag{7}$$

The structure of the irreducible representations of the r.h.s. of Eq. (7) is (the upper labels denote different irreducible representations):

$$(\tfrac{3}{2},\tfrac{3}{2}) \equiv \begin{pmatrix} \mathrm{CCC} & \mathrm{UCC} & \mathrm{UUC} & \mathrm{UUU} \\ \mathrm{GCC} & \mathrm{ACC} & \mathrm{AUC} & \mathrm{AUU} \\ \mathrm{GGC} & \mathrm{AGC} & \mathrm{AAC} & \mathrm{AAU} \\ \mathrm{GGG} & \mathrm{AGG} & \mathrm{AAG} & \mathrm{AAA} \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2})^1 \equiv \begin{pmatrix} \mathrm{CCG} & \mathrm{UCG} & \mathrm{UUG} & \mathrm{UUA} \\ \mathrm{GCG} & \mathrm{ACG} & \mathrm{AUG} & \mathrm{AUA} \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2})^2 \equiv \begin{pmatrix} \mathrm{CGC} & \mathrm{UGC} & \mathrm{UAC} & \mathrm{UAU} \\ \mathrm{CGG} & \mathrm{UGG} & \mathrm{UAG} & \mathrm{UAA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{3}{2})^1 \equiv \begin{pmatrix} \mathrm{CCU} & \mathrm{UCU} \\ \mathrm{GCU} & \mathrm{ACU} \\ \mathrm{GGU} & \mathrm{AGU} \\ \mathrm{GGA} & \mathrm{AGA} \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{3}{2})^2 \equiv \begin{pmatrix} \mathrm{CUC} & \mathrm{CUU} \\ \mathrm{GUC} & \mathrm{GUU} \\ \mathrm{GAC} & \mathrm{GAU} \\ \mathrm{GAG} & \mathrm{GAA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})^1 \equiv \begin{pmatrix} \mathrm{CCA} & \mathrm{UCA} \\ \mathrm{GCA} & \mathrm{ACA} \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{1}{2})^2 \equiv \begin{pmatrix} \mathrm{CGU} & \mathrm{UGU} \\ \mathrm{CGA} & \mathrm{UGA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})^3 \equiv \begin{pmatrix} \mathrm{CUG} & \mathrm{CUA} \\ \mathrm{GUG} & \mathrm{GUA} \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{1}{2})^4 \equiv \begin{pmatrix} \mathrm{CAC} & \mathrm{CAU} \\ \mathrm{CAG} & \mathrm{CAA} \end{pmatrix}$$

The correspondence with the amino-acids is given in Table 10 (for the eukariotic code).

Let us close this section by drawing the reader's attention to Fig. 1 where is specified for each codon its position in the appropriate representation. The diagram of states for each representation is supposed to lie in a separate parallel plane. Thick lines connect codons associated to the same amino-acid. One remarks that each segment relates a couple of codons belonging to the same representation or to two different representations. This last case occurs for quadruplets or sextets of codons associated to the same amino-acid.

3 The Reading (or Ribosome) operator $\mathcal{R}$

3.1 General structure of the reading operator

As expected from formula (7), our model does not gather codons associated to one particular amino-acid in the same irreducible multiplet. However, it is possible to construct an operator $\mathcal{R}$ out of the algebra $\mathcal{U}_{q\to 0}(sl(2)\oplus sl(2))$, acting on the codons, that will describe the various genetic codes in the following way:

Two codons have the same eigenvalue under $\mathcal{R}$ if and only if they are associated to the same amino-acid. This operator $\mathcal{R}$ will be called the reading operator.

It is a remarkable fact that the various genetic codes share the same basic structure. As we mentioned above, the dinucleotides can be split into "strong" dinucleotides CC, GC, UC, AC, CU, GU, CG and GG that lead to quartets and "weak" ones UU, AU, UG, AG, CA, GA, UA, AA that lead to doublets. Let us construct a prototype of the reading operator that reproduces this structure.

The first part of the reading operator $\mathcal{R}$ is responsible for the structure in quadruplets given essentially by the dinucleotide content. It is given by (the c_i are arbitrary coefficients)

$$\tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3}\ . \tag{8}$$

The operators $J_{\alpha,3}$ ($\alpha = H, V$) are the third components of the total spin generators of the algebra $\mathcal{U}_{q\to 0}(sl(2)\oplus sl(2))$. The operator C_α is a Casimir operator of $\mathcal{U}_{q\to 0}(sl(2)_\alpha)$ in the crystal basis. It commutes with $J_{\alpha\pm}$ and $J_{\alpha,3}$ and its eigenvalues on any vector basis of an irreducible representation of highest weight J is $J(J+1)$, that is the same as the undeformed standard second degree Casimir operator of $sl(2)$. Its explicit expression is

$$C_\alpha = (J_{\alpha,3})^2 + \tfrac{1}{2}\sum_{n\in\mathbb{Z}_+}\sum_{k=0}^{n}(J_{\alpha-})^{n-k}(J_{\alpha+})^n(J_{\alpha-})^k\ . \tag{9}$$

Note that for $sl(2)_{q\to 0}$ the Casimir operator is an infinite series of powers of $J_{\alpha\pm}$. However in any finite irreducible representation only a finite number of terms gives a non-vanishing contribution.

$\mathcal{P}_H$ and $\mathcal{P}_V$ are projectors given by the following expressions:

$$\mathcal{P}_H = J^d_{H+}\, J^d_{H-} \qquad \text{and} \qquad \mathcal{P}_V = J^d_{V+}\, J^d_{V-} \; . \tag{10}$$

The second part of $\mathcal{R}$ gives rise to the splitting of the quadruplets into doublets. It reads

$$-\, 2\mathcal{P}_D\, c_3\, J_{V,3} \tag{11}$$

where the projector $\mathcal{P}_D$ is given by

$$\begin{aligned} \mathcal{P}_D \;=\; & (1 - J^d_{V+}\, J^d_{V-})(J^d_{H+}\, J^d_{H-})(J^d_{H-}\, J^d_{H+}) + (1 - J^d_{H+}\, J^d_{H-})(1 - J^d_{V+}\, J^d_{V-}) \\ & + (1 - J^d_{H+}\, J^d_{H-})(J^d_{V+}\, J^d_{V-})(J^d_{H-}\, J^d_{H+}) \; . \end{aligned} \tag{12}$$

The third part of $\mathcal{R}$ allows to reproduce the sextets viewed as quartets plus doublets. It is

$$-\, 2\mathcal{P}_S\, c_4\, J_{V,3} \tag{13}$$

where the projector $\mathcal{P}_S$ is given by

$$\mathcal{P}_S = (J^d_{H-}\, J^d_{H+})\; [(J^d_{H+}\, J^d_{H-})(1 - J^d_{V+}\, J^d_{V-}) + (J^d_{V+}\, J^d_{V-})(J^d_{V-}\, J^d_{V+})(1 - J^d_{H+}\, J^d_{H-})] \; . \tag{14}$$

At this point, one obtains the eigenvalues of the reading operator $\mathcal{R}$ for the 64 codons, where Y = C,U (pyrimidines), R = G,A (purines) and N = C,U,G,A:

$$\begin{array}{ll}
\mathrm{CCN} = -c_1 - c_2 & \mathrm{GCN} = -c_1 + 3c_2 \\
\mathrm{UCN} = 3c_1 - c_2 & \mathrm{ACN} = 3c_1 + 3c_2 \\
\mathrm{CUN} = c_1 - c_2 & \mathrm{GUN} = c_1 + 3c_2 \\
\mathrm{CGN} = -c_1 + c_2 & \mathrm{GGN} = -c_1 + 5c_2 \\
\mathrm{UUY} = 5c_1 - c_2 - 3c_3 & \mathrm{UUR} = 5c_1 - c_2 - c_3 \\
\mathrm{AUY} = 5c_1 + 3c_2 - c_3 - c_4 & \mathrm{AUR} = 5c_1 + 3c_2 + c_3 + c_4 \\
\mathrm{UGY} = 3c_1 + c_2 - c_3 - c_4 & \mathrm{UGR} = 3c_1 + c_2 + c_3 + c_4 \\
\mathrm{AGY} = 3c_1 + 5c_2 + c_3 + c_4 & \mathrm{AGR} = 3c_1 + 5c_2 + 3c_3 + 3c_4 \\
\mathrm{CAY} = c_1 + c_2 - c_3 & \mathrm{CAR} = c_1 + c_2 + c_3 \\
\mathrm{GAY} = c_1 + 5c_2 + c_3 & \mathrm{GAR} = c_1 + 5c_2 + 3c_3 \\
\mathrm{UAY} = 5c_1 + c_2 - c_3 & \mathrm{UAR} = 5c_1 + c_2 + c_3 \\
\mathrm{AAY} = 5c_1 + 5c_2 + c_3 & \mathrm{AAR} = 5c_1 + 5c_2 + 3c_3
\end{array} \tag{15}$$

The coefficients c_3 and c_4 are fixed as follows. The coefficient c_3 is set to the value $c_3 = 4c_1$ by requiring that the quartet CUN and the doublet UUR, associated to the amino-acid `Leu`, lead to the same $\mathcal{R}$-eigenvalue. It remains to reproduce the `Ser` sextet. This is achieved by taking for the coefficient c_4 the value $c_4 = -4c_1 - 6c_2$, such that the final eigenvalues for the codons are the following:

$$\begin{array}{llll}
\mathrm{CCN} = -c_1 - c_2 & \mathrm{GCN} = -c_1 + 3c_2 & \mathrm{UCN} = 3c_1 - c_2 & \mathrm{ACN} = 3c_1 + 3c_2 \\
\mathrm{CUN} = c_1 - c_2 & \mathrm{GUN} = c_1 + 3c_2 & \mathrm{CGN} = c_1 + c_2 & \mathrm{GGN} = -c_1 + 5c_2 \\
\mathrm{UUY} = -7c_1 - c_2 & \mathrm{UUR} = c_1 - c_2 & \mathrm{AUY} = 5c_1 + 9c_2 & \mathrm{AUR} = 5c_1 - 3c_2 \\
\mathrm{UGY} = 3c_1 + 7c_2 & \mathrm{UGR} = 3c_1 - 5c_2 & \mathrm{AGY} = 3c_1 - c_2 & \mathrm{AGR} = 3c_1 - 13c_2 \\
\mathrm{CAY} = -3c_1 + c_2 & \mathrm{CAR} = 5c_1 + c_2 & \mathrm{GAY} = 5c_1 + 5c_2 & \mathrm{GAR} = 13c_1 + 5c_2 \\
\mathrm{UAY} = c_1 + c_2 & \mathrm{UAR} = 9c_1 + c_2 & \mathrm{AAY} = 9c_1 + 5c_2 & \mathrm{AAR} = 17c_1 + 5c_2
\end{array} \tag{16}$$

The prototype of the reading operator $\mathcal{R}$ takes finally the form:

$$\mathcal{R} = \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + (-8c_1\, \mathcal{P}_D\, + (8c_1 + 12c_2)\, \mathcal{P}_S)\, J_{V,3} \tag{17}$$

and the correspondence codons/amino-acids is given as follows:

$$\begin{array}{llll}
\mathrm{CCN} \to \texttt{Pro} & \mathrm{UCN} \to \texttt{Ser} & \mathrm{GCN} \to \texttt{Ala} & \mathrm{ACN} \to \texttt{Thr} \\
\mathrm{CUN} \to \texttt{Leu} & \mathrm{GUN} \to \texttt{Val} & \mathrm{CGN} \to \texttt{Arg} & \mathrm{GGN} \to \texttt{Gly} \\
\mathrm{UUY} \to \texttt{Phe} & \mathrm{AUY} \to \texttt{Ile} & \mathrm{UGY} \to \texttt{Cys} & \mathrm{AGY} \to \texttt{Ser} \\
\mathrm{UUR} \to \texttt{Leu} & \mathrm{AUR} \to \texttt{Met} & \mathrm{UGR} \to \texttt{Trp} & \mathrm{AGR} \to \text{unassigned (X)} \\
\mathrm{CAY} \to \texttt{His} & \mathrm{UAY} \to \texttt{Tyr} & \mathrm{GAY} \to \texttt{Gln} & \mathrm{AAY} \to \texttt{Asn} \\
\mathrm{CAR} \to \texttt{Gln} & \mathrm{UAR} \to \texttt{Ter} & \mathrm{GAR} \to \texttt{Glu} & \mathrm{AAR} \to \texttt{Lys}
\end{array} \tag{18}$$

3.2 The various genetic codes

In this section, we will determine the reading operators for the following genetic codes:
– the Eukariotic Code (EC),
– the Vertebral Mitochondrial Code (VMC),
– the Yeast Mitochondrial Code (YMC),
– the Invertebrate Mitochondrial Code (IMC),
– the Protozoan Mitochondrial and Mycoplasma Code (PMC),
– the Echinoderm Mitochondrial Code (EMC),
– the Ascidian Mitochondrial Code (AMC),
– the Flatworm Mitochondrial Code (FMC),
– the Ciliate Nuclear Code (CNC),
– the Blepharisma Nuclear Code (BNC),
– the Euplotid Nuclear Code (ENC),
– the Alternative Yeast Nuclear Code (alt. YNC),
Let us emphasize that each of these codes is very close to the assignment (18). The main differences between the biological codes and the prototype code (18) are the following:

- assignment of the doublet AGR either to Arg (codes EC, YMC, PMC, CNC, BNC, ENC, aYNC), Ser (codes IMC, EMC, FMC), Gly (code AMC) or the stop signal Ter (code VMC).

  Such an assignment is done by the following term in the reading operator:

$$c_5\, \mathcal{P}_{AG} \left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) \tag{19}$$

The operators $J^{(3)}_{\alpha,3}$ are the third components corresponding to the third nucleotide of a codon. Of course, these last two operators can be replaced by $J^{(3)}_{\alpha,3} = J_{\alpha,3} - J^{d}_{\alpha,3}$.
The projector $\mathcal{P}_{AG}$ is given by

$$\mathcal{P}_{AG} = (J^d_{H+}\, J^d_{H-})(J^d_{H-}\, J^d_{H+})(1 - J^d_{V+}\, J^d_{V-})(J^d_{V-}\, J^d_{V+}) \tag{20}$$

and the coefficient c_5 by

$$\begin{array}{ll} \text{for } \texttt{Arg} & c_5 = -4c_1 + 14c_2 \\ \text{for } \texttt{Ser} & c_5 = 12c_2 \\ \text{for } \texttt{Gly} & c_5 = -4c_1 + 18c_2 \\ \text{for } \texttt{Ter} & c_5 = 6c_1 + 14c_2 \end{array} \tag{21}$$

- splitting of some doublets into singlets (one element of the singlet combining to another doublet to form a triplet):
  Met → Met + Ile for the EC, PMC, EMC, FMC, CNC, BNC, ENC, aYNC codes;
  Lys → Lys + Asn for the FMC and EMC codes;
  Trp → Trp + Ter for the EC, CNC, BNC, aYNC codes;
  Trp → Trp + Cys for the ENC code;
  Ter → Tyr + Ter for the FMC code;
  Such an assignment is done through the following term in the reading operator:

$$c_6\, \mathcal{P}_{XY} \left(\tfrac{1}{2} - J^{(3)}_{V,3}\right)\left(\tfrac{1}{2} - J^{(3)}_{H,3}\right) \tag{22}$$

where we use the projector $\mathcal{P}_{AU}$ for the splitting of the Met doublet, $\mathcal{P}_{AA}$ for the Lys doublet, $\mathcal{P}_{UG}$ for the Trp doublet, and $\mathcal{P}_{UA}$ for the Ter doublet. These projectors are given by

$$\mathcal{P}_{AU} = (1 - J^d_{H+}\, J^d_{H-})(J^d_{H-}\, J^d_{H+})(J^d_{V+}\, J^d_{V-})(J^d_{V-}\, J^d_{V+}) \tag{23}$$

$$\mathcal{P}_{AA} = (1 - J^d_{H+}\, J^d_{H-})(J^d_{H-}\, J^d_{H+})(1 - J^d_{V+}\, J^d_{V-})(J^d_{V-}\, J^d_{V+}) \tag{24}$$

$$\mathcal{P}_{UG} = (J^d_{H+}\, J^d_{H-})(J^d_{H-}\, J^d_{H+})(1 - J^d_{V+}\, J^d_{V-})(1 - J^d_{V-}\, J^d_{V+}) \tag{25}$$

$$\mathcal{P}_{UA} = (1 - J^d_{H+}\, J^d_{H-})(J^d_{H-}\, J^d_{H+})(1 - J^d_{V+}\, J^d_{V-})(1 - J^d_{V-}\, J^d_{V+}) \tag{26}$$

The coefficient c_6 takes the following values:

$$\begin{array}{ll} \text{for } \texttt{Met} \longrightarrow \texttt{Met} + \texttt{Ile} & c_6 = 12c_2 \\ \text{for } \texttt{Lys} \longrightarrow \texttt{Lys} + \texttt{Asn} & c_6 = -8c_1 \\ \text{for } \texttt{Trp} \longrightarrow \texttt{Trp} + \texttt{Cys} & c_6 = 12c_2 \\ \text{for } \texttt{Trp} \longrightarrow \texttt{Trp} + \texttt{Ter} & c_6 = 6c_1 + 6c_2 \\ \text{for } \texttt{Ter} \longrightarrow \texttt{Ter} + \texttt{Tyr} & c_6 = -8c_1 \end{array} \tag{27}$$

- in the case of the CNC and BNC codes, the Ter doublet is changed in Gln as follows:
  Ter → Gln for the CNC code by the term

$$-\,4c_1\, \mathcal{P}_{UA} \left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) \tag{28}$$

  Ter → Ter + Gln for the BNC code by the term

$$-\,4c_1\, \mathcal{P}_{UA} \left(\tfrac{1}{2} - J^{(3)}_{V,3}\right)\left(\tfrac{1}{2} + J^{(3)}_{H,3}\right) \tag{29}$$

- in the case of the alternative YNC code, the last quartet Leu is split into a triplet Leu coded by (CUC,CUU,CUA) and a doublet Ser coded by (CUG). The corresponding term in the reading operator is

$$2c_1\,\mathcal{P}_{CU}\left(\tfrac{1}{2}-J^{(3)}_{V,3}\right)\left(\tfrac{1}{2}+J^{(3)}_{H,3}\right) \tag{30}$$

where the projector $\mathcal{P}_{CU}$ is given by

$$\mathcal{P}_{CU}=(1-J^d_{H+}\,J^d_{H-})(1-J^d_{H-}\,J^d_{H+})(J^d_{V+}\,J^d_{V-})(1-J^d_{V-}\,J^d_{V+}) \tag{31}$$

- in the case of the Yeast Mitochondrial Code, the quartet CUN codes the amino-acid Thr rather than Leu. This change is achieved by multiplying the quartets term (8) by $(1+2\mathcal{P}_{CU})$ for the horizontal part and by $(1-4\mathcal{P}_{CU})$ for the vertical part.

3.2.1 The Eukariotic Code (EC)

The Eukariotic Code is the most important one and is often referred to as the universal code. The differences between the Eukariotic Code and the prototype code are the following:

	prototype code	EC		prototype code	EC
AUG	Met	Met	AUA	Met	Ile
AGG	X	Arg	AGA	X	Arg
UGG	Trp	Trp	UGA	Trp	Ter

Hence from (19), (21), (22) and (27), the reading operator for the Eukariotic Code is

$$\begin{aligned}\mathcal{R}_{EC} \;=\; & \tfrac{4}{3}c_1\,C_H+\tfrac{4}{3}c_2\,C_V-4c_1\,\mathcal{P}_H\,J_{H,3}-4c_2\,\mathcal{P}_V\,J_{V,3}+(-8c_1\,\mathcal{P}_D\,+(8c_1+12c_2)\,\mathcal{P}_S)\,J_{V,3}\\ & +(-4c_1+14c_2)\,\mathcal{P}_{AG}\left(\tfrac{1}{2}-J^{(3)}_{V,3}\right)\\ & +\Big[12c_2\,\mathcal{P}_{AU}+(6c_1+6c_2)\,\mathcal{P}_{UG}\Big]\left(\tfrac{1}{2}-J^{(3)}_{V,3}\right)\left(\tfrac{1}{2}-J^{(3)}_{H,3}\right)\end{aligned} \tag{32}$$

3.2.2 The Vertebral Mitochondrial Code (VMC)

The Vertebral Mitochondrial Code is used in the mitochondriae of vertebrata. The differences between the Vertebral Mitochondrial Code and the prototype code are the following:

	prototype code	VMC		prototype code	VMC
AGG	X	Ter	AGA	X	Ter

Hence from (19) and (21), the reading operator for the Vertebral Mitochondrial Code is

$$\begin{aligned}\mathcal{R}_{VMC} \;=\; & \tfrac{4}{3}c_1\,C_H+\tfrac{4}{3}c_2\,C_V-4c_1\,\mathcal{P}_H\,J_{H,3}-4c_2\,\mathcal{P}_V\,J_{V,3}+(-8c_1\,\mathcal{P}_D\,+(8c_1+12c_2)\,\mathcal{P}_S)\,J_{V,3}\\ & +(6c_1+14c_2)\,\mathcal{P}_{AG}\left(\tfrac{1}{2}-J^{(3)}_{V,3}\right)\end{aligned} \tag{33}$$

3.2.3 The Yeast Mitochondrial Code (YMC)

The Yeast Mitochondrial Code is used in the mitochondriae of yeast such as Saccharomyces, Candida, etc. The differences between the Yeast Mitochondrial Code and the prototype code are the following:

	prototype code	YMC		prototype code	YMC
CUC	`Leu`	`Thr`	CUU	`Leu`	`Thr`
CUG	`Leu`	`Thr`	CUA	`Leu`	`Thr`
AGG	X	`Arg`	AGA	X	`Arg`

Hence from (19) and (21), the reading operator for the Yeast Mitochondrial Code is

$$\begin{aligned} \mathcal{R}_{YMC} &= (\tfrac{4}{3}c_1\, C_H - 4c_1\, \mathcal{P}_H\, J_{H,3})(1+2\mathcal{P}_{CU}) + (\tfrac{4}{3}c_2\, C_V - 4c_2\, \mathcal{P}_V\, J_{V,3})(1-4\mathcal{P}_{CU}) \\ &\quad +(-8c_1\, \mathcal{P}_D\ + (8c_1+12c_2)\, \mathcal{P}_S)\, J_{V,3} + (-4c_1+14c_2)\, \mathcal{P}_{AG}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) \end{aligned} \tag{34}$$

3.2.4 The Invertebrate Mitochondrial Code (IMC)

The Invertebrate Mitochondrial Code is used in the mitochondriae of some arthopoda, mollusca, nematoda and insecta. The differences between the Invertebrate Mitochondrial Code and the prototype code are the following:

	prototype code	IMC		prototype code	IMC
AGG	X	`Ser`	AGA	X	`Ser`

Hence from (19) and (21), the reading operator for the Invertebrate Mitochondrial Code is

$$\begin{aligned} \mathcal{R}_{IMC} &= \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + (-8c_1\, \mathcal{P}_D\ + (8c_1+12c_2)\, \mathcal{P}_S)\, J_{V,3} \\ &\quad +12c_2\, \mathcal{P}_{AG}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) \end{aligned} \tag{35}$$

3.2.5 The Protozoan Mitochondrial and Mycoplasma Code (PMC)

The Protozoan Mitochondrial and Mycoplasma Code is used in the mitochondriae of some protozoa (leishmania, paramecia, trypanosoma, etc.) and for many fungi. The differences between the Protozoan Mitochondrial and Mycoplasma Code and the prototype code are the following:

	prototype code	PMC		prototype code	PMC
AUG	`Met`	`Met`	AUA	`Met`	`Ile`
AGG	X	`Arg`	AGA	X	`Arg`

Hence from (19), (21), (22) and (27), the reading operator for the Protozoan Mitochondrial Code is

$$\begin{aligned} \mathcal{R}_{PMC} &= \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + (-8c_1\, \mathcal{P}_D\ + (8c_1+12c_2)\, \mathcal{P}_S)\, J_{V,3} \\ &\quad +(-4c_1+14c_2)\, \mathcal{P}_{AG}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) + 12c_2\, \mathcal{P}_{AU}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right)\left(\tfrac{1}{2} - J^{(3)}_{H,3}\right) \end{aligned} \tag{36}$$

3.2.6 The Echinoderm Mitochondrial Code (EMC)

The Echinoderm Mitochondrial Code is used in the mitochondriae of some asterozoa and echinozoa. The differences between the Echinoderm Mitochondrial Code and the prototype code are the following:

	prototype code	EMC		prototype code	EMC
AUG	Met	Met	AUA	Met	Ile
AGG	X	Ser	AGA	X	Ser
AAG	Lys	Lys	AAA	Lys	Asn

Hence from (19), (21), (22) and (27), the reading operator for the Echinoderm Mitochondrial Code is

$$\begin{aligned}\mathcal{R}_{EMC} \;=\;& \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + \left(-8c_1\, \mathcal{P}_D + (8c_1 + 12c_2)\, \mathcal{P}_S\right) J_{V,3} \\ & +12c_2\, \mathcal{P}_{AG}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) + \Big[12c_2\, \mathcal{P}_{AU} - 8c_1\, \mathcal{P}_{AA}\Big]\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right)\left(\tfrac{1}{2} - J^{(3)}_{H,3}\right) \qquad (37)\end{aligned}$$

3.2.7 The Ascidian Mitochondrial Code (AMC)

The Ascidian Mitochondrial Code is used in the mitochondriae of some ascidiacea. The differences between the Ascidian Mitochondrial Code and the prototype code are the following:

	prototype code	AMC		prototype code	AMC
AGG	X	Gly	AGA	X	Gly

Hence from (19) and (21), the reading operator for the Ascidian Mitochondrial Code is

$$\begin{aligned}\mathcal{R}_{AMC} \;=\;& \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + \left(-8c_1\, \mathcal{P}_D + (8c_1 + 12c_2)\, \mathcal{P}_S\right) J_{V,3} \\ & +(-4c_1 + 18c_2)\, \mathcal{P}_{AG}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) \qquad (38)\end{aligned}$$

3.2.8 The Flatworm Mitochondrial Code (FMC)

The Flatworm Mitochondrial Code is used in the mitochondriae of the flatworms. The differences between the Flatworm Mitochondrial Code and the prototype code are the following:

	prototype code	FMC		prototype code	FMC
UAG	Ter	Ter	UAA	Ter	Tyr
AUG	Met	Met	AUA	Met	Ile
AGG	X	Ser	AGA	X	Ser
AAG	Lys	Lys	AAA	Lys	Asn

Hence from (19), (21), (22) and (27), the reading operator for the Flatworm Mitochondrial Code is

$$\begin{aligned}\mathcal{R}_{FMC} \;=\;& \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + \left(-8c_1\, \mathcal{P}_D + (8c_1 + 12c_2)\, \mathcal{P}_S\right) J_{V,3} \\ & +12c_2\, \mathcal{P}_{AG}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) + \Big[12c_2\, \mathcal{P}_{AU} - 8c_1\, \mathcal{P}_{AA} - 8c_1\, \mathcal{P}_{UA}\Big]\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right)\left(\tfrac{1}{2} - J^{(3)}_{H,3}\right) \\ & \qquad (39)\end{aligned}$$

3.2.9 The Ciliate Nuclear Code (CNC)

The Ciliate Nuclear Code is used in the nuclei of some ciliata, dasyclasaceae and diplomonadida. The differences between the Ciliate Nuclear Code and the prototype code are the following:

	prototype code	CNC		prototype code	CNC
UGG	Trp	Trp	UGA	Trp	Ter
UAG	Ter	Gln	UAA	Ter	Gln
AUG	Met	Met	AUA	Met	Ile
AGG	X	Arg	AGA	X	Arg

Hence from (19), (21), (22), (27) and (28), the reading operator for the Ciliate Nuclear Code is

$$\begin{aligned}\mathcal{R}_{CNC} \;=\;& \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + (-8c_1\, \mathcal{P}_D\, + (8c_1 + 12c_2)\, \mathcal{P}_S)\, J_{V,3} \\ & + \Big[(-4c_1 + 14c_2)\, \mathcal{P}_{AG} - 4c_1\, \mathcal{P}_{UA}\Big]\Big(\tfrac{1}{2} - J^{(3)}_{V,3}\Big) \\ & + \Big[12c_2\, \mathcal{P}_{AU} + (6c_1 + 6c_2)\, \mathcal{P}_{UG}\Big]\Big(\tfrac{1}{2} - J^{(3)}_{V,3}\Big)\Big(\tfrac{1}{2} - J^{(3)}_{H,3}\Big) \qquad (40)\end{aligned}$$

3.2.10 The Blepharisma Nuclear Code (BNC)

The Blepharisma Nuclear Code is used in the nuclei of the blepharisma (ciliata) (note that this code is very close to the CNC which is used for the ciliata). The differences between the Blepharisma Nuclear Code and the prototype code are the following:

	prototype code	BNC		prototype code	BNC
UGG	Trp	Trp	UGA	Trp	Ter
UAG	Ter	Gln	UAA	Ter	Ter
AUG	Met	Met	AUA	Met	Ile
AGG	X	Arg	AGA	X	Arg

Hence from (19), (21), (22), (27) and (29), the reading operator for the Blepharisma Nuclear Code is

$$\begin{aligned}\mathcal{R}_{BNC} \;=\;& \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + (-8c_1\, \mathcal{P}_D\, + (8c_1 + 12c_2)\, \mathcal{P}_S)\, J_{V,3} \\ & + (-4c_1 + 14c_2)\, \mathcal{P}_{AG}\, \Big(\tfrac{1}{2} - J^{(3)}_{V,3}\Big) - 4c_1\, \mathcal{P}_{UA}\, \Big(\tfrac{1}{2} - J^{(3)}_{V,3}\Big)\Big(\tfrac{1}{2} + J^{(3)}_{H,3}\Big) \\ & + \Big[12c_2\, \mathcal{P}_{AU} + (6c_1 + 6c_2)\, \mathcal{P}_{UG}\Big]\Big(\tfrac{1}{2} - J^{(3)}_{V,3}\Big)\Big(\tfrac{1}{2} - J^{(3)}_{H,3}\Big) \qquad (41)\end{aligned}$$

3.2.11 The Euplotid Nuclear Code (ENC)

The Euplotid Nuclear Code is used in the nuclei of the euplotidae (ciliata). The differences between the Euplotid Nuclear Code and the prototype code are the following:

	prototype code	ENC		prototype code	ENC
UGG	Trp	Trp	UGA	Trp	Cys
AUG	Met	Met	AUA	Met	Ile
AGG	X	Arg	AGA	X	Arg

Hence from (19), (21), (22) and (27), the reading operator for the Euplotid Nuclear Code is

$$\begin{aligned}\mathcal{R}_{ENC} \;=\;& \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + (-8c_1\, \mathcal{P}_D\, + (8c_1 + 12c_2)\, \mathcal{P}_S)\, J_{V,3} \\ & + (-4c_1 + 14c_2)\, \mathcal{P}_{AG}\, \Big(\tfrac{1}{2} - J^{(3)}_{V,3}\Big) + 12c_2\, (\mathcal{P}_{AU} + \mathcal{P}_{UG})\Big(\tfrac{1}{2} - J^{(3)}_{V,3}\Big)\Big(\tfrac{1}{2} - J^{(3)}_{H,3}\Big) \qquad (42)\end{aligned}$$

3.2.12 The alternative Yeast Nuclear Code (alt. YNC)

The alternative Yeast Nuclear Code is used in the nuclei of some yeast (essentially many candidae). The differences between the alternative Yeast Nuclear Code and the prototype code are the following:

	prototype code	alt. YNC		prototype code	alt. YNC
CUG	Leu	Ser	CUA	Leu	Leu
UGG	Trp	Trp	UGA	Trp	Ter
AUG	Met	Met	AUA	Met	Ile
AGG	X	Arg	AGA	X	Arg

Hence from (19), (21), (22), (27) and (30), the reading operator for the alternative Yeast Nuclear Code is

$$\begin{aligned}\mathcal{R}_{aYNC} \;=\;& \tfrac{4}{3}c_1\, C_H + \tfrac{4}{3}c_2\, C_V - 4c_1\, \mathcal{P}_H\, J_{H,3} - 4c_2\, \mathcal{P}_V\, J_{V,3} + (-8c_1\, \mathcal{P}_D\, + (8c_1 + 12c_2)\, \mathcal{P}_S)\, J_{V,3} \\ & +(-4c_1 + 14c_2)\, \mathcal{P}_{AG}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right) + 2c_1\, \mathcal{P}_{CU}\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right)\left(\tfrac{1}{2} + J^{(3)}_{H,3}\right) \\ & +\Big[(6c_1 + 6c_2)\, \mathcal{P}_{UG} + 12c_2\, \mathcal{P}_{AU}\Big]\left(\tfrac{1}{2} - J^{(3)}_{V,3}\right)\left(\tfrac{1}{2} - J^{(3)}_{H,3}\right) \qquad (43)\end{aligned}$$

3.3 Reading values for the amino-acids

We have therefore constructed reading operators for the genetic codes specified above, starting from a prototype code that emphasizes the quartet/doublet structure of the different codes. The different reading operators are such that they give the same value for a given amino-acid, whatever the code under consideration. Finally, we get the following eigenvalues of the reading operators for the amino-acids (after a rescaling, setting $c \equiv c_1/c_2$):

a.a.	value of $\mathcal{R}$	a.a.	value of $\mathcal{R}$	a.a.	value of $\mathcal{R}$
Ala	$-c+3$	Gly	$-c+5$	Pro	$-c-1$
Arg	$-c+1$	His	$-3c+1$	Ser	$3c-1$
Asn	$9c+5$	Ile	$5c+9$	Thr	$3c+3$
Asp	$5c+5$	Leu	$c-1$	Trp	$3c-5$
Cys	$3c+7$	Lys	$17c+5$	Tyr	$c+1$
Gln	$5c+1$	Met	$5c-3$	Val	$c+3$
Glu	$13c+5$	Phe	$-7c-1$	Ter	$9c+1$

(44)

Remark that the reading operators $\mathcal{R}(c)$ can be used for any real value of c, except those conferring the same eigenvalue to codons relative to two different amino-acids. These forbidden values are the following: -7, -5, -4, -3, $-\frac{5}{2}$, $-\frac{7}{3}$, -2, $-\frac{5}{3}$, $-\frac{3}{2}$, $-\frac{4}{3}$, -1, $-\frac{5}{6}$, $-\frac{4}{5}$, $-\frac{3}{4}$, $-\frac{5}{7}$, $-\frac{2}{3}$, $-\frac{3}{5}$, $-\frac{1}{2}$, $-\frac{3}{7}$, $-\frac{2}{5}$, $-\frac{3}{8}$, $-\frac{1}{3}$, $-\frac{3}{10}$, $-\frac{2}{7}$, $-\frac{1}{4}$, $-\frac{2}{9}$, $-\frac{1}{5}$, $-\frac{1}{6}$, $-\frac{1}{7}$, $-\frac{1}{8}$, $-\frac{1}{9}$, 0, $\frac{1}{7}$, $\frac{1}{6}$, $\frac{1}{5}$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{2}{5}$, $\frac{1}{2}$, $\frac{2}{3}$, 1, $\frac{4}{3}$, $\frac{3}{2}$, 2, $\frac{5}{2}$, 3, 4, 5.

At this point, let us emphasize the specific properties of our model. To each nucleotide are assigned specific quantum numbers characterizing its purine/pyrimidine origin and involving the complementary rule. Then ordered sequences of bases can be constructed and characterized in this framework. Ordered sequences of three bases have been just above examined and the correspondence codon/amino-acid represented by the reading operator $\mathcal{R}$. Finally let us remark that the coefficients c_i, which above have been taken as constants, can more generally be considered as functions of some external variables (biological, physical and chemical environment, time, etc.). In this way it is possible to explain the observed discrepancy in the correspondence codons/amino-acid in biological species under stress conditions (in vitro). In this scheme the evolution process of genetic code can also be discussed. However, we believe that a better understanding of the reasons of the evolution, i.e. which kind of optimization process takes place, has still to be acquired.

4 Physical properties of the dinucleotides

The model we have at hand, with nucleotides characterized by quantum numbers, is well adapted to elaborate formulae expressing biophysical properties. A particularly interesting quantity is the free energy released by base pairing in double stranded RNA. The data are not provided for a doublet of nucleotides, with one item in each strand, but for a pair of nucleotides, for ex. CG, lying on one strand and coupled with another pair, i.e. GC on the second strand ; note also that the direction on a strand being perfectly defined, the release of energy for the doublet sequence CG on the first strand running from $5'$ to $3'$ related to the doublet GC on the complementary strand running from $3'$ to $5'$, will be different to the one related to the doublet GC, itself associated to CG. It appears clear that such quantities involve pairs of nucleotides, and that naturally ordered crystal bases obtained from tensor product of two representations are adapted for such a calculation.

We will also consider two other quantities involving again pairs of nucleotides, namely the relative hydrophilicity R_f and hydrophobicity R_x of dinucleosides.

Before presenting our results, let us mention that fits for the same biophysical properties can be found in a recent preprint [5] where polynomials in 4 or 6 coordinates in the 64 codon space are constructed. In their approach, the authors associate two coordinates (d, m) to each nucleotide of any codon, as follows: $A = (-1, 0)$, $C = (0, -1)$, $G = (0, 1)$, $U = (1, 0)$, labelling in this way each codon with 6 numbers. The above labelling of the nucleotides is related to our labels Eq. (2) in the following way:

	d	m
C	$J_{V,3} - J_{H,3}$	$-(J_{V,3} + J_{H,3})$
U	$J_{V,3} - J_{H,3}$	$-(J_{V,3} + J_{H,3})$
G	$J_{V,3} + J_{H,3}$	$J_{V,3} - J_{H,3}$
A	$J_{V,3} + J_{H,3}$	$J_{V,3} - J_{H,3}$

Therefore the labels (d, m) just correspond up to a sign for the pyrimidine (resp. purine) to the antidiagonal and diagonal (resp. diagonal and antidiagonal) $\mathcal{U}_{q\to 0}(sl(2))$.

In the following we compare our results with those of [5].

Free energy

In [1] we have fitted the experimental data with a four-parameter operator. Here we fit the more recent data [6] with a two-parameter operator obtained from the one used in [1] by setting two parameters to zero:

$$\Delta G^0_{37} = \alpha_0 + \alpha_1 (C_H + C_V) J^d_{3H} \tag{45}$$

Using a least-squares fit, one finds for the coefficients α_i:

$$\alpha_0 = -2.14\,, \quad \alpha_1 = -0.295 \tag{46}$$

The standard deviation of the two-parameter fit (46) is found to be equal to 0.149, which is to be compared to the standard deviation 0.16 of the four-parameter fit of ref. [5]. The experimental and fitted values of the free energies ΔG^0_{37} of the dinucleotides are displayed in Table 1.

CA	−2.1 −2.14	CG	−2.4 −2.73	UG	−2.1 −2.14	UA	−1.3 −1.55
CU	−2.1 −2.14	CC	−3.3 −3.32	UC	−2.4 −2.14	UU	−0.9 −0.96
GU	−2.2 −2.14	GC	−3.4 −3.32	AC	−2.2 −2.14	AU	−1.1 −0.96
GA	−2.4 −2.14	GG	−3.3 −3.32	AG	−2.1 −2.14	AA	−0.9 −0.96

Table 1: Dinucleotides free energies ΔG^0_{37}.
The upper (resp. lower) values are the experimental (resp. fitted) values.

Hydrophilicity

We fit the values of the relative hydrophilicity R_f of the 16 dinucleoside monophosphates [7] with the following four-parameter operator:

$$R_f = \alpha_0 + \alpha_1 C_V + \alpha_2 J^d_{3V} + \alpha_3 \sum_{i=1,2} (J^i_{3H} + J^i_{3V})(J^i_{3H} + J^i_{3V} - 1) \tag{47}$$

(the last term in α_3 is equal to 4 for AA, to 2 for CA, GA, UA and zero for the other dinucleotides).
Using a least-squares fit, one finds for the coefficients α_i:

$$\alpha_0 = 0.135\,, \quad \alpha_1 = 0.036\,, \quad \alpha_2 = 0.147\,, \quad \alpha_3 = -0.016 \tag{48}$$

The standard deviation of the four-parameter fit (48) is found to be equal to 0.027, which is to be compared to the standard deviation 0.033 of the six-parameter fit of ref. [5]. The experimental and fitted values of the hydrophilicity R_f of the dinucleosides are displayed in Table 2.

$$\begin{bmatrix}\mathrm{CA} \begin{matrix}0.083\\0.103\end{matrix}\end{bmatrix} \qquad \begin{bmatrix}\mathrm{CG} \begin{matrix}0.146\\0.135\end{matrix} & \mathrm{UG} \begin{matrix}0.160\\0.135\end{matrix} & \mathrm{UA} \begin{matrix}0.090\\0.103\end{matrix}\end{bmatrix}$$

$$\begin{bmatrix}\mathrm{CU} \begin{matrix}0.359\\0.354\end{matrix}\\ \\ \mathrm{GU} \begin{matrix}0.224\\0.207\end{matrix}\\ \\ \mathrm{GA} \begin{matrix}0.035\\0.028\end{matrix}\end{bmatrix} \quad \begin{bmatrix}\mathrm{CC} \begin{matrix}0.349\\0.354\end{matrix} & \mathrm{UC} \begin{matrix}0.378\\0.354\end{matrix} & \mathrm{UU} \begin{matrix}0.389\\0.354\end{matrix}\\ & & \\ \mathrm{GC} \begin{matrix}0.193\\0.207\end{matrix} & \mathrm{AC} \begin{matrix}0.118\\0.175\end{matrix} & \mathrm{AU} \begin{matrix}0.112\\0.175\end{matrix}\\ & & \\ \mathrm{GG} \begin{matrix}0.065\\0.060\end{matrix} & \mathrm{AG} \begin{matrix}0.048\\0.028\end{matrix} & \mathrm{AA} \begin{matrix}0.023\\-0.004\end{matrix}\end{bmatrix}$$

Table 2: Dinucleosides relative hydrophilicities R_f.
The upper (resp. lower) values are the experimental (resp. fitted) values.

Hydrophobicity

We fit the values of the relative hydrophobicity R_x of the 16 dinucleoside monophosphates as reported in [8] with the following four-parameter operator:

$$R_x = \alpha_0 + \alpha_1 J^d_{3V} + \alpha_2 J^d_{3H} + \alpha_3[(J^1_{3H} + J^1_{3V})^2 + (J^2_{3H} + J^2_{3V})^2] \tag{49}$$

(the last term in α_3 is equal to 2 for AA, AC, CA and CC, to 2 for AU, AG, UA, UC, GC, GA, CU and CG and zero for UU, UG, GU, GG).
Using a least-squares fit, one finds for the coefficients α_i:

$$\alpha_0 = 0.294\,, \quad \alpha_1 = -0.240\,, \quad \alpha_2 = -0.105\,, \quad \alpha_3 = 0.136 \tag{50}$$

Using a least-squares fit without the dinucleoside AA, one finds new coefficients α_i, which lead to better values of R_x for the remaining dinucleosides:

$$\alpha_0 = 0.309\,, \quad \alpha_1 = -0.203\,, \quad \alpha_2 = -0.068\,, \quad \alpha_3 = 0.099 \tag{51}$$

The standard deviation of the four-parameter fit (50) is equal to 0.049, which is the same of the four-parameter fit of ref. [5]. Using the fit (51), the standard deviation becomes 0.074

$$\begin{bmatrix}\mathrm{CA} \begin{matrix}0.494\\0.507\end{matrix}\end{bmatrix} \qquad \begin{bmatrix}\mathrm{CG} \begin{matrix}0.326\\0.340\end{matrix} & \mathrm{UG} \begin{matrix}0.291\\0.309\end{matrix} & \mathrm{UA} \begin{matrix}0.441\\0.476\end{matrix}\end{bmatrix}$$

$$\begin{bmatrix}\mathrm{CU} \begin{matrix}0.218\\0.205\end{matrix}\\ \\ \mathrm{GU} \begin{matrix}0.291\\0.309\end{matrix}\\ \\ \mathrm{GA} \begin{matrix}0.660\\0.611\end{matrix}\end{bmatrix} \quad \begin{bmatrix}\mathrm{CC} \begin{matrix}0.244\\0.236\end{matrix} & \mathrm{UC} \begin{matrix}0.218\\0.205\end{matrix} & \mathrm{UU} \begin{matrix}0.194\\0.174\end{matrix}\\ & & \\ \mathrm{GC} \begin{matrix}0.326\\0.340\end{matrix} & \mathrm{AC} \begin{matrix}0.494\\0.507\end{matrix} & \mathrm{AU} \begin{matrix}0.441\\0.476\end{matrix}\\ & & \\ \mathrm{GG} \begin{matrix}0.436\\0.444\end{matrix} & \mathrm{AG} \begin{matrix}0.660\\0.611\end{matrix} & \mathrm{AA} \begin{matrix}1\\0.778\end{matrix}\end{bmatrix}$$

Table 3: Dinucleosides relative hydrophobicities R_x.
The upper (resp. lower) values are the experimental (resp. fitted) values.

(including the value for AA) or 0.024 (excluding the value for AA). For this last case, the standard deviation of ref. [5] is still equal to 0.031. The experimental and fitted values (second fit) of the relative hydrophobicity R_x of the dinucleosides are displayed in Table 3.

5 Universal behaviour of ratios of codon usage frequency

In the following the labels X, J, Z, K represent any of the 4 bases C, U, G, A. Let XJZ be a codon in a given multiplet, say m_i, encoding an a.a., say A_i. We define the probability of usage of the codon XJZ as the ratio between the frequency of usage n_Z of the codon XJZ in the biosynthesis of A_i and the total number n of synthesized A_i, i.e. as the relative codon frequency, in the limit of *very large* n.

It is natural to assume that the usage frequency of a codon in a multiplet is connected to its probability of usage $P(XJZ \to A_i)$. We define [2] the *branching ratio* B_{ZK} as

$$B_{ZK} = \frac{P(XJZ \to A_i)}{P(XJK \to A_i)} \tag{52}$$

where XJK is another codon belonging to the same multiplet m_i. It is reasonable to argue that in the limit of very large number of codons, for a fixed biological species and amino-acid, the branching ratio depends essentially on the properties of the codon. In our model this means that in this limit B_{ZK} is a function, depending on the type of the multiplet, on the *quantum numbers* of the codons XJZ and XJK, i.e. on the labels $J_\alpha, J_{\alpha,3}$, where $\alpha = H$ or V, and on an other set of quantum labels leaving out the degeneracy on J_α; in Table 4 different irreducible representations with the same values of J_α are distinguished by an upper label.

We have put in evidence a correlation in the codon usage frequency for the quartets and the quartet subpart of the sextets, i.e. the codons in a sextet differing only for the third codon, for the vertebrates in [2] and for biological species belonging to the vertebrates, invertebrates, plants and fungi in [9], and we have shown that these correlations fit well in our model with the assumed dependence on B_{ZK}. Here we remark that for thirteen biological species belonging to the vertebrate class, with a statistics of codons larger than 95,000 (see Table 5), the ratio of

$$\frac{B_{AG}}{B_{UC}} = \frac{B_{AU}}{B_{GC}} = \frac{P(XJA \to A_i)}{P(XJG \to A_i)} \frac{P(XJC \to A_i)}{P(XJU \to A_i)} \tag{53}$$

for quartets and the quartet subpart of the sextets has a behaviour independent of the specific biological species. Moreover, for the same amino-acids for which we have remarked correlations, the values of the ratio B_{AG}/B_{UC} are almost the same (see Table 8). We show that these behaviour and correlations find a nice explanation in our model. In Tables 6 and 7, we report respectively the values of the branching ratios B_{AG} and B_{UC} as computed from the database [10] (release of February 2000) and in Table 8 the ratio of these quantities. The average values $\langle B_{AG}/B_{UC} \rangle$, the standard deviations σ and the ratios $\sigma / \langle B_{AG}/B_{UC} \rangle$ are displayed in the following table:

	Pro	Ala	Thr	Ser	Gly	Val	Leu	Arg
$\langle B_{AG}/B_{UC}\rangle$	2.50	2.84	3.30	2.67	2.21	0.33	0.26	1.32
σ	0.46	0.53	0.56	0.35	0.30	0.04	0.03	0.14
$\sigma/\langle B_{AG}/B_{UC}\rangle$	0.19	0.19	0.17	0.13	0.14	0.13	0.10	0.11

The above behaviour can be easily understood considering a dependence on B_{ZK} not only on the irreducible representations to which the codons XJZ and XJK appearing in the numerator and the denominator belong, but also on the specific states denoting these codons, and refining the factorized form of [2] as

$$B_{ZK} = F_{ZK}(IR(XJZ); IR(XJK))\,\frac{G_H(b.s.; J_{H,3}(XJZ))\,G_V(b.s.; J_{V,3}(XJZ))}{G_H(b.s.; J_{H,3}(XJK))\,G_V(b.s.; J_{H,3}(XJK))} \tag{54}$$

where we have denoted by $b.s.$ the biological species, by $IR(XJZ)$ and $J_{\alpha,3}(XJZ)$ the irreducible representation to which the codon XJZ belong (see Table 4), and the value of the third component of the α-spin of the state XJZ. Note that we have still neglected the dependence on the type of the biosynthetized amino-acid. *The ratio B_{AG}/B_{UC} using Eq. (54), is no more depending on the biological species but only on the value of the irreducible representations of the codons.* Moreover, for `Pro`, `Ala`, `Thr`, `Ser`, (resp. `Val` and `Leu`), the irreducible representations appearing in the F functions are the same as can be seen from Table 9, so we expect the same value for the ratio, which is indeed the case (see above Table), the value of B_{AG}/B_{UC} for the first four amino-acids (resp. for the last two amino-acids) lying in the range $2.90 \pm 15\%$ (resp. $0.30 \pm 15\%$). These values should be compared with the value 1.32 for `Arg` and 2.21 for `Gly`.

Let us end this section by the following remark. From the above table, one might be tempted to consider the value of the ratio B_{AG}/B_{UC} for `Gly` of the same order of magnitude as the ones for `Pro`, `Ala`, `Thr`, `Ser`. Then one distinguishes, following this ratio, three groups of codons quartets: the one associated to the five just mentioned amino-acids, another one relative to `Val` and `Leu`, and a last one with `Arg`. Now, let us look at the dinucleotide pairs constituting the first two nucleotides in a codon in the light of our results of sect. 2: the pairs CC, GC, AC, UC and GG relative to `Pro`, `Ala`, `Thr`, `Ser`, and `Gly` respectively belong to the representation (1,1) of $\mathcal{U}_{q\to 0}(sl(2) \oplus sl(2))$; the states GU and CU relative to `Val` and `Leu` respectively belong to the representation (0,1); finally CG relative to `Arg` also lies in a different representation (1,0).

6 Mutations in the genetic code

In this section, we present a mathematical framework to describe the single-base deletions in the genetic code. In [11] starting from the observation that the single-base deletions in DNA, which occur far more frequently that single base additions, take place in the opposite site to a purine **R**, (**R = G, A**) i.e. a pyrimidine **Y** (**Y = C, U/T**) is deleted, arguments have been presented to explain why the Stop codons have the structure they have, see Table 4. We refer to the paper for more details and for references to the biological literature on the subject and we recall here just the main ideas and conclusions of [11]. The starting point is the observed fact that deletions occur more frequently in the following sequences: **YR**, **TTR**, **YTG** and **TR**. In ref. [11] all these sequences have been refined as **YTRV**, (**V = C, A, G**). Starting from the structure of this dangerous sequence and using the complementarity property, an analysis shows that four codons – **TAA**, **TAG**, **TTA**, **CTA** – are both potential deletion site codons and reverse-complementary potential site codons. As a mutation at the end of a protein chain just implies the addition of further peptides, the authors conclude that the assignment of codons **TAA** and **TAG** as Stop codons minimizes the possible deleterious effects of deletion. Indeed the codon usage frequency of the dangerous codon **CTA**, as it can be seen from fig. (5) of [2] and from fig. (2) of [9], is very low. An analysis of the codon usage frequency exhibits an analogous behaviour for the codon **TTA**.

The mechanism by which the above specified sequences are preferred in the deletion process is unclear. In the following we will present a mathematical scheme in which these properties can be settled. Let us recall that the Wigner-Eckart theorem, has been extended to the quantum algebra $U_q(sl(n))$, and recently in [12] to the case of $U_{q\to 0}(sl(2))$.

In [12] ($q \to 0$)-tensor operators have been introduced, called crystal tensor operators, which transform as

$$J_3(\tau^j_m) \equiv m\tau^j_m \quad J_\pm(\tau^j_m) \equiv \tau^j_{m\pm 1} \tag{55}$$

Clearly, if $|m| > j$ then τ^j_m has to be considered vanishing.

The ($q \to 0$)-Wigner-Eckart theorem can be written ($j_1 \geq j$)

$$\begin{aligned} \tau^j_m\,|j_1 m_1\rangle &= (-1)^{2j} \sum_{\alpha=0}^{2j} \langle j_1 + j - \alpha \| \tau^j \| j_1 \rangle \, |j_1 + j - \alpha, m_1 + m\rangle \\ & \quad (\delta_{m_1, j_1-\alpha} + \delta_{-m, j-\alpha} - \delta_{m_1, j_1-\alpha}\,\delta_{-m, j-\alpha}) \end{aligned} \tag{56}$$

The ($q \to 0$)-Wigner-Eckart theorem has the peculiar feature that the selection rules do not depend only on the rank of the tensor operator and on the initial state, but in a crucial way from the specific component of the tensor in consideration. The tensor product of two irreducible representations in the crystal basis is not commutative (see sect. 2), therefore one has to specify which is the first representation. In the following, as in [12], the crystal tensor operator has to be considered as the first one.

Let us also remark the following peculiar property of crystal basis which will be used in the following. We specify it only for the case we are interested in, but it is a completely general property.

An ordered sequence, or chain, of n nucleotides is a state belonging to an irreducible representation of $U_{q\to 0}((sl(2) \oplus sl(2))$ appearing in the n-fold product of the fundamental irreducible representation $(1/2, 1/2)$. Moreover the same property holds for any subsequence of m ($m < n$) nucleotides. We can mimick the deletion of a N nucleotide in a generic position of a coding sequence by a local annihilation operator of the N nucleotide. In order to take into account the observed fact that the deletion of the nucleotide depends on the nature of the neighboring nucleotides, we require the annihilation operator to behave as a defined crystal tensor operator under $U_{q\to 0}(sl(2))_V$ or $U_{q\to 0}(sl(2))_H$ or both. In our mathematical description we have to specify the action of the annihilation operator on a chain of nucleotides. If we assume that the annihilation of the N nucleotide behaves e.g. as a spinor crystal operator for the $U_{q\to 0}(sl(2))_V$, we have to require that the deletion of the N nucleotide from the initial chain of K nucleotides, described by the state $|J_i, M_i; \Omega_i\rangle$, leading to the final chain of $K-1$ nucleotides, described by the state $|J_f, M_f; \Omega_f\rangle$, is compatible with the $(q \to 0)$-Wigner-Eckart theorem prescription for the action of the definite crystal spinor operator between the initial state $|J_i, M_i; \Omega_i\rangle$ and the final state $|J_f, M_f; \Omega_f\rangle$, where we have denoted by Ω the set of all the labels necessary to identify completely the state. As we shall see, this is far from being trivial and will put constraints on the type of nucleotides surrounding the nucleotide N. We have to specify which chain has to be considered in order to study the action of the crystal tensor operator. It seems reasonable to take into account chains formed by $K = 2$ and 3 nucleotides starting from N in the sense of the reading of the codon sequence. So we are defining on the chain the action of a "matrioska" crystal tensor operator. We assume:

Assumption : The biological mechanism responsible for the deletion of a pyrimidine **C** (resp. **U**) in a sequence can be schematized by a local crystal tensor operator $\tau^{1/2}_{-1/2}$ for $U_{q\to 0}(sl(2)_V)$ and $\tau^{1/2}_{-1/2}$ (resp. $\tau^{1/2}_{1/2}$) for $U_{q\to 0}(sl(2)_H)$, which transforms the state **YX** (resp. **YXZ**) into the state **X** (resp. **XZ**), **X**, **Z** being any nucleotide.

By "local crystal tensor operator" we mean an operator which, in the sequence of RNA, acts on the K-chain ($K = 2, 3$) starting with **Y**, deleting the pyrimidine, according to the selection rules imposed by the assumed type of the crystal tensor.

Let us point out that, differently to ref. [11], where the DNA sequence was analyzed, we consider the transcripted RNA sequence and the deletion in the trascription of a **Y**.

There are 8 possible cases (we denote the initial and final states with the notation of sect. 2 and by A (resp. F) the allowed (resp. forbidden) transition). We analyze the deletion of a **C** (on the left) and of an **U** (on the right).

Action of $\tau^{1/2}_{-1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(1,1) \to (\frac{1}{2},\frac{1}{2})$		
CC	**C**	F–F
$(0,1) \to (\frac{1}{2},\frac{1}{2})$		
CU	**U**	A–F
$(1,0) \to (\frac{1}{2},\frac{1}{2})$		
CG	**G**	F–A
$(0,0) \to (\frac{1}{2},\frac{1}{2})$		
CA	**A**	A–A

Action of $\tau^{1/2}_{1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(1,1) \to (\frac{1}{2},\frac{1}{2})$		
UC **UU**	**C** **U**	A–F A–F
$(1,0) \to (\frac{1}{2},\frac{1}{2})$		
UG **UA**	**G** **A**	A–A A–A

So for the transition for the state of dinucleotide to one nucleotide state, from the assumed nature of the crystal tensor operator, it follows that a pyrimidine can be deleted if followed by a purine. Now let us consider what happens if we consider the transition from a trinucleotide to a dinucletide state. Using the previous result we consider only the state in which a purine is in second position so we have to consider 16 cases:

Action of $\tau^{1/2}_{-1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(\frac{1}{2}, \frac{1}{2})^4 \to (1,1)$		
CAC	**AC**	A–A
CAU	**AU**	A–A
CAA	**AA**	A–A
CAG	**AG**	A–A
$(\frac{3}{2}, \frac{1}{2})^2 \to (1,1)$		
CGC	**GC**	F–A
CGG	**GG**	F–A
$(\frac{1}{2}, \frac{1}{2})^2 \to (0,1)$		
CGU	**GU**	F–A
CGA	**GA**	F–A

Action of $\tau^{1/2}_{1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(\frac{3}{2}, \frac{1}{2})^2 \to (1,1)$		
UAC	**AC**	A–A
UAU	**AU**	A–A
UAA	**AA**	A–A
UAG	**AG**	A–A
$(\frac{3}{2}, \frac{1}{2})^2 \to (1,1)$		
UGC	**GC**	A–A
UGG	**GG**	A–A
$(\frac{1}{2}, \frac{1}{2})^2 \to (0,1)$		
UGU	**GU**	A–A
UGA	**GA**	A–A

So, from the assumed nature of the crystal tensor operator, the transition from a trinucleotide to a dinucleotide state is horizontally forbidden for the deletion of a **C** if the second nucleotide is a **G**.

Let us note that we have made the simplified assuption that the transitions depend only on the values of J_α, $J_{\alpha,3}$ of the initial and final state.

Moreover, both to take into account the data of [11] and to check that the results are not very sensible to the choice of the initial state, we consider the deletion of a purine in second position in a four-nucleotide state and impose that the process may take place only if the initial and final state can be connected by a spinor crystal operator $\tau^{1/2}_{-1/2,H} \oplus \tau^{1/2}_{-1/2,V}$ for the deletion of **C** or $\tau^{1/2}_{1/2,H} \oplus \tau^{1/2}_{-1/2,V}$ for the deletion of **U**.

As the two pyrimidines differ by their value of $J_{H,3}$, the constraints imposed by the tensor operator $\tau^{1/2}_{\pm 1/2,H}$ are weaker than those imposed by the tensor operator $\tau^{1/2}_{-1/2,V}$.

In Appendix (in sect. 2) we have reported all the irreducible representations arising by the 4-fold (3-fold) tensor product of the fundamental representation. A detailed analysis shows that only the following deletions may happen (we report all the transitions that are allowed at least once):

Action of $\tau^{1/2}_{-1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(2,1)^3 \to (\frac{3}{2},\frac{3}{2})$		
GCGC	**GGC**	F–A
ACGC	**AGC**	F–A
GCGG	**GGG**	F–A
ACGG	**AGG**	F–A
$(2,0)^2 \to (\frac{3}{2},\frac{1}{2})^2$		
CCGG	**CGG**	F–A
UCGG	**UGG**	F–A
$(1,1)^7 \to (\frac{1}{2},\frac{3}{2})^1$		
GCGU	**GGU**	F–A
ACGU	**AGU**	F–A
GCGA	**GGA**	F–A
ACGA	**AGA**	F–A
$(1,0)^4 \to (\frac{1}{2},\frac{1}{2})^2$		
CCGA	**CGA**	F–A
UCGA	**UGA**	F–A
$(1,1)^9 \to (\frac{1}{2},\frac{3}{2})^2$		
GCAC	**GAC**	F–A
GCAG	**GAG**	F–A
$(1,0)^6 \to (\frac{1}{2},\frac{1}{2})^4$		
CCAG	**CAG**	F–A
$(1,1)^9 \to (\frac{3}{2},\frac{3}{2})$		
ACAC	**AAC**	A–A
ACAU	**AAU**	A–A
ACAG	**AAG**	A–A
ACAA	**AAA**	A–A
$(1,0)^6 \to (\frac{3}{2},\frac{1}{2})^2$		
UCAG	**UAG**	A–A
UCAA	**UAA**	A–A

Action of $\tau^{1/2}_{-1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(1,2)^3 \to (\frac{3}{2},\frac{3}{2})$		
UCUC	**UUC**	A–F
UCUU	**UUU**	A–F
ACUC	**AUC**	A–F
ACUU	**AUU**	A–F
$(0,2)^2 \to (\frac{1}{2},\frac{3}{2})^2$		
CCUU	**CUU**	A–F
GCUU	**GUU**	A–F
$(1,1)^8 \to (\frac{3}{2},\frac{1}{2})^1$		
UCUG	**UUG**	A–F
UCUA	**UUA**	A–F
ACUG	**AUG**	A–F
ACUA	**AUA**	A–F
$(0,1)^5 \to (\frac{1}{2},\frac{1}{2})^3$		
CCUA	**CUA**	A–F
GCUA	**GUA**	A–F
$(1,1)^9 \to (\frac{3}{2},\frac{1}{2})^2$		
UCAC	**UAC**	A–F
UCAU	**UAU**	A–F
$(0,1)^6 \to (\frac{1}{2},\frac{1}{2})^4$		
CCAU	**CAU**	A–F
$(0,0)^4 \to (\frac{1}{2},\frac{1}{2})^4$		
CCAA	**CAA**	A–A
$(0,1)^6 \to (\frac{1}{2},\frac{3}{2})^2$		
GCAU	**GAU**	A–A
GCAA	**GAA**	A–A

So we remark:

- The deletion of **C**, allowed or horizontally forbidden, may happen only if it is followed by a purine. In the allowed cases, it must be followed by the nucleotide **A**, in agreement with the observed data.

- A nucleotide **A** before the deleted nucleotide **C** appears only in the transition $(1,1)^9 \to (\frac{3}{2},\frac{3}{2})$. This feature is present in the observed data with a very low occurrence, which in our language would mean that the matrix element of τ between these two irreducible representations is small.

Now we consider the case of deletion of **U**. A detailed analysis shows that only the following deletions may happen:

Action of $\tau^{1/2}_{1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(1,2)^1 \to (\frac{1}{2},\frac{3}{2})^2$		
CUUC	**CUC**	A–F
CUUU	**CUU**	A–F
GUUC	**GUC**	A–F
GUUU	**GUU**	A–F
$(1,2)^2 \to (\frac{3}{2},\frac{3}{2})$		
CUCC	**CCC**	A–F
GUCC	**GCC**	A–F
$(1,1)^3 \to (\frac{1}{2},\frac{1}{2})^4$		
CUAC	**CAC**	A–F
CUAU	**CAU**	A–F
$(1,1)^6 \to (\frac{1}{2},\frac{1}{2})^3$		
UUCA	**UCA**	A–F
AUCA	**ACA**	A–F
$(2,1)^2 \to (\frac{3}{2},\frac{1}{2})^1$		
UUCG	**UCG**	A–F
UUUG	**UUG**	A–F
UUUA	**UUA**	A–F
AUCG	**ACG**	A–F
AUUG	**AUG**	A–F
AUUA	**AUA**	A–F
$(2,1)^3 \to (\frac{3}{2},\frac{1}{2})^2$		
UUGC	**UGC**	A–F
UUAC	**UAC**	A–F
UUAU	**UAU**	A–F
$(1,1)^7 \to (\frac{1}{2},\frac{1}{2})^3$		
UUGU	**UGU**	A–F
$(0,1)^4 \to (\frac{1}{2},\frac{1}{2})^2$		
CUGU	**CGU**	A–F

Action of $\tau^{1/2}_{1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(1,1)^2 \to (\frac{1}{2},\frac{1}{2})^3$		
CUUG	**CUG**	A–F
GUUG	**GUG**	A–F
CUUA	**CUA**	A–F
GUUA	**GUA**	A–F
$(1,1)^2 \to (\frac{3}{2},\frac{1}{2})^1$		
CUCG	**CCG**	A–F
GUCG	**GCG**	A–F
$(1,2)^2 \to (\frac{1}{2},\frac{3}{2})^1$		
UUCU	**UCU**	A–F
AUCU	**ACU**	A–F
$(0,2)^1 \to (\frac{1}{2},\frac{3}{2})^1$		
CUCU	**CCU**	A–F
GUCU	**GCU**	A–F
$(2,2) \to (\frac{3}{2},\frac{3}{2})$		
UUCC	**UCC**	A–F
UUUC	**UUC**	A–F
UUUU	**UUU**	A–F
AUCC	**ACC**	A–F
AUUC	**AUC**	A–F
AUUU	**AUU**	A–F
$(0,1)^3 \to (\frac{1}{2},\frac{1}{2})^1$		
CUCA	**CCA**	A–F
GUCA	**GCA**	A–F
$(1,1)^3 \to (\frac{3}{2},\frac{1}{2})^2$		
CUGC	**CGC**	A–F

Action of $\tau^{1/2}_{1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(1,1)^7 \to (\frac{1}{2},\frac{3}{2})^1$		
AUGU	**AGU**	A–A
AUGA	**AGA**	A–A
$(0,1)^4 \to (\frac{1}{2},\frac{3}{2})^1$		
GUGU	**GGU**	A–A
GUGA	**GGA**	A–A
$(1,1)^3 \to (\frac{1}{2},\frac{3}{2})^2$		
GUAC	**GAC**	A–A
GUAU	**GAU**	A–A
GUAG	**GAG**	A–A
GUAA	**GAA**	A–A
$(2,0)^2 \to (\frac{3}{2},\frac{1}{2})^2$		
UUGG	**UGG**	A–A
UUAG	**UAG**	A–A
UUAA	**UAA**	A–A
$(1,0)^2 \to (\frac{3}{2},\frac{1}{2})^2$		
CUGG	**CGG**	A–A

Action of $\tau^{1/2}_{1/2,H} \oplus \tau^{1/2}_{-1/2,V}$

$(1,1)^3 \to (\frac{3}{2},\frac{3}{2})$		
GUGC	**GGC**	A–A
GUGG	**GGG**	A–A
$(1,0)^2 \to (\frac{1}{2},\frac{1}{2})^4$		
CUAG	**CAG**	A–A
CUAA	**CAA**	A–A
$(2,1)^3 \to (\frac{3}{2},\frac{3}{2})$		
AUGC	**AGC**	A–A
AUAC	**AAC**	A–A
AUAU	**AAU**	A–A
AUGG	**AGG**	A–A
AUAG	**AAG**	A–A
AUAA	**AAA**	A–A
$(0,0)^2 \to (\frac{1}{2},\frac{1}{2})^2$		
CUGA	**CGA**	A–A

So we remark:

- The deletion of **U** may happen only if it is followed by **A** or by **G**. In the observed data only **A** is considered; however in [11] the reported deletion of **U** are about 1/4 with respect to the reported deletion of **C**. So our modelisation just foresees a different environment for the deletion of **U** and **C**.
- The last nucleotide in the four-nucleotide sequence in which the deletion occurs may be any nucleotide, but the case in which it is a purine seems more frequent than the case in which it is a pyrimidine.
- There are no transition which are only horizontally forbidden.

In conclusion, both from considering the transitions on the K-chains ($K = 2,3$) to the $(K-1)$-chains or the transition from the four-nucleotide states to the three-nucleotide states under the action of the crystal tensor operators, we deduce that the deletion of a pyrimidine may happen if it is followed by a purine. In particular, for the deletion of C the preferred purine is the adenine A, whilst for the deletion of U also the guanine G may appear. This makes a difference between the two cases and it would be extremely interesting to see if more accurate data may confirm this asymmetry. Moreover the next following nucleotide may be of any type but there is indication that a purine is preferred. So our mathematical scheme explains the main features of the observed data [11]. A more quantitative analysis should require higher statistics in the experimental data.

7 Recent theoretical approaches: a comparison

The use of continuous symmetries in the genetic code has been considered by different teams these recent years[2]. It appears of some importance to summarize each of these approaches, and to make clear how the model we propose differ from them.

In 1993, an underlying symmetry based on a continuous group has been proposed [13]. More precisely, considering the eukaryotic code, the authors tried to answer the following question: is it possible to determine a Lie algebra $\mathcal{G}$ carrying a 64-dimensional irreducible representation R and admitting a subalgebra $\mathcal{H}$ such that the decomposition of R into irreducible multiplets under $\mathcal{H}$ gives exactly the 21 different multiplets, the different codons in each of the first 20 multiplets being associated to the same amino-acid, the last multiplet containing the stop codons ? They proposed as starting symmetry the symplectic algebra $sp(6)$, which indeed admits an irreducible representation of dimension 64, equal to the number of different codons, with the successive breakings:

$$sp(6) \supset sp(4) \oplus su(2) \supset su(2) \oplus su(2) \oplus su(2) \supset su(2) \oplus U(1) \oplus su(2) \supset su(2) \oplus U(1) \oplus U(1) \quad (57)$$

Such a chain of symmetry breaking could be considered as reflecting the evolution of the genetic code, the six amino-acids relative to the codons in the irreducible representations obtained after the first breaking (in which $64 = 16 + 4 + 20 + 10 + 12 + 2$) appearing as primordial amino-acids in their approach. However, the authors were obliged, in order to reproduce the actual multiplet pattern, to assume in the final breaking, a partial breaking or a "freezing" in the sense that the breaking of the last $su(2)$ into $U(1)$ does not occur for all the multiplets. As an example, such a freezing has to be imposed to the sextets corresponding to Leu and Ser, which otherwise would decompose into three doublets. In the same way, freezing will forbid the doublets related to Lys and Cys to split into singlets.

In a second further paper, dated 1997 [14], a refinement of this approach has been considered, with the use of Lie groups instead of Lie algebras: then, global properties, for example non connexity of $O(2) = U(1) \times \mathbb{Z}_2$, can be exploited. In this context, the authors proposed another chain of breaking starting with the exceptional group G_2, which also allows a 64 dimensional irreducible representation. But here again, the freezing pathology cannot be avoided.

One can also mention the work of [15] where the unifying algebra before breaking is $so(14)$.

Meantime (1997), interpreting the double origin of the nucleotides, each arising either from purine or from pyrimidine, as a $\mathbb{Z}_2$-grading a supersymmetric model was proposed [16], involving superalgebras for such a program. The $\mathbb{Z}_2$-grading specific of a simple superalgebra is there used to separate purine and pyrimidine: indeed, by putting the four nucleotids in the the 4 dimensional representation of $su(2/1)$ one can confer to the A and G purines (R) an even grading, and to the C and U pyrimidines (Y) an odd grading; note that the R states are then in the $su(2)$ doublet and the Y ones $su(2)$ singlets. The notion of polarity spin is also introduced,

[2] See section "Symmetry techniques in Biological Systems" in Proc. XXII Int. Coll. on Group Theoretical Methods in Physics, pp. 142-165.

allowing to distinguish the C and G nucleotides with two locally polarized sites, from the A and U ones with three polarized sites: the C and G (resp. A and U) will be assigned in a doublet (resp. in singlets) of another $su(2)$. Then the authors consider the sum of algebras: $su(2) \oplus su(2) \oplus su(2|1)$ with the first (second) $su(2)$ acting as polarity spin on the first (second) nucleotid of a codon, and the $su(2|1)$ acting on the third nucleotid only. Moreover the two $su(2)$ would act in an alternating way on the first and second position, that is as $1/2$, $-1/2$ and $-1/2$, $1/2$. This sum of algebras can be embedded in the superalgebra $su(6|1)$, which admits a 64 dimensional irreducible representation, and could be also used for a superalgebraic approach to the genetic code evolution, with the chain of symmetry breaking:

$$\begin{aligned} su(6|1) \supset su(2) \oplus su(3|1) \supset su(2) \oplus su(2) \oplus su(2|1) \supset U(1) \oplus U(1) \oplus su(2|1) \\ \supset U(1) \oplus U(1) \oplus gl(1|1) \end{aligned} \tag{58}$$

Again the problem of freezing, that is the last breaking applies to some but not all the multiplets, is present with this choice of (super)algebras.

It seems necessary to remark that in this proposal which implies (super)algebras acting in the same time on nucleotides and on codons – one must say in a rather complicated way – the nucleotides cannot appear as building blocks from which one algebraically constructs the codons, by performing tensorial products of representations, as is the case of our model. In fact, the problem of ordering the nucleotides inside a codon forbids this natural way of proceeding as long as only usual (super)algebras are involved. Note that it is the limit of quantum algebras that we use in our approach: then, we have at hand the so-called crystal bases, which exactly solve the ordering problem.

In a last month preprint, two authors of the same team [5] proposed to fit biophysical properties of nucleic acids by constructing polynomials in 6 coordinates in the 64 dimensional codon space. As already mentioned in sect. 4, the two coordinates they associate to each nucleotide is direcetly related to the nucleotide eigenvalues of our model. The authors present their computations as independent of a particular choice of algebra or superalgebra as long as the underlying algebra is of rank 6 – which is in particular the dimension of the Cartan subalgebra of $su(6|1)$ – and admits a 64 dimensional irreducible representation. We note that our model does allow to calculate the biophysical quantities considered in ref. [5] without the constraint on representations, but more importantly, with only a two rank algebra.

A detailed and systematic study of superalgebras and superalgebra breaking chains has been performed by the authors of [17]: it is the orthosymplectic $osp(5|2)$ superalgebra which emerges from their algebraic analysis.

Finally, it is amazing to remark that, just a few years after the the concept of genetic code was formulated, an attempt to give a mathematical description of its properties was started by the russian physicist Yu. B. Rumer [18]. Indeed he remarked that the 16 *roots*, i.e. the combinations of the first two codons, divide in a *strong octet* which form quartets ou sub-part of sextets and a *weak octet* which form doublets, triplets and singlets, attempting to give a systematic description of the genetic code. A few years after, with B.G. Konopel'chenko [19] they formulated the strong assumption that with respect to any property of the codons the 16 roots can be gathered into two octets with opposite "charge", whose positive (negative) value respectively characterizes the strong and weak roots. This description comes out naturally in our model, such a charge Q being defined in Eq. (5) of sect. 2.

8 Conclusion

Our model is based on the algebra $\mathcal{U}_{q\to0}(sl(2)\oplus sl(2))$ that we have chosen for two main characteristics. First it encodes the stereochemical property of a base, and also reflects the complementarity rule, by conferring quantum numbers to each nucleotide. Secondly, it admits representation spaces or crystal bases in which an ordered sequence of nucleotides or codon can be suitably characterized. Let us emphasize that $\mathcal{U}_{q\to0}(sl(2)\oplus sl(2))$ is really neither a Lie algebra nor an enveloping deformed algebra. We still use in a loose sense the word algebra, just to emphasize the fact that we use largely the mathematical tools of representation space, tensor operators etc. which are typical of the algebraic structures. Let us add that it is a remarkable property of a quantum algebra in the limit $q\to0$ to admit representations, obtained from the tensorial product of basic ones, in which each state appears as a unique sequence of ordered basic elements.

In this framework, the correspondence codon/amino-acid is realized by the operator $\mathcal{R}_c$, constructed out of the symmetry algebra, and acting on codons: the eigenvalues provided by $\mathcal{R}_c$ on two codons will be equal or different depending on whether the two codons are associated to the same or to two different amino-acids. It is remarkable that this correspondence can be obtained for all the genetic codes and that the reading operators have a bulk common to the various genetic codes (the prototype reading operator) and differ only for a few additive terms, analogous to perturbative terms present in most Hamiltonians describing complex physical systems. Moreover they depend on parameters, presently assumed as constants, which in principle can be considered as functions of suitable variables. These feature may be of some interest in the study of the evolution of the genetic code, problem which has not yet been tackled in our model.

Then, restricting to the case of states made of two nucleotides, the experimental values of the free energy, released by base pairing in the formation of double stranded nucleic acids, of the hydrophibicity and of the hydrophilicity have been fitted with expressions depending respectively on 2, 4 and 4 parameters and constructed out of the generators of $\mathcal{U}_{q\to0}(sl(2)\oplus sl(2))$.

The model does not necessarily assign the codons in a multiplet (in particular the quartets, sextets and triplet) to the same irreducible representation. Let us remark that the assignments of the codons to the different irreducible representations is a straightforward consequence of the tensor product, once assigned the nucleotides to the fundamental irreducible representation. This feature is relevant, since it can explain the correlation between the branching ratios of the codon usage of different codons coding the same amino-acid as discussed in [2] and [9]. Here we have shown that the universal pattern (inside the class of vertebrates) of B_{AG}/B_{UC} can simply be reproduced in our model.

Moreover our mathematical description of the genetic code allows a modelisation of some biological process. A first step in this direction has been presented in sect. 6, where we have shown that the observed data related to the a pyrimidine deletion can be simulated by introducing the concept of $q \to 0$ – or crystal – tensor operator. Finally let us mention some directions for future development of our model. Going further in the analysis of the branching ratios, we want to refine our analysis and make a more detailed study taking into account the dependence on the family of biological species. Indeed preliminary analysis on plants, invertebrates and bacteriae shows that, even if the pattern of the correlation is still approximatively present, large deviations appear which presumably exhibit evidence that the dependence on subclass or family of biological species cannot any more be neglected, differently to the case of vertebrates. A further investigation of the possibility of mathematically modelising or simulating biological processes, in particular mutations, by crystal tensor operators, is in progress. Other questions are still to be investigated: in particular how could the genetic code evolution be reproduced in our model ?

References

[1] L. Frappat, A. Sciarrino, P. Sorba, *A crystal base for the genetic code,* Phys. Lett. A **250** (1998) 214 and `physics/9801027`.

[2] L. Frappat, A. Sciarrino, P. Sorba, *Symmetry and codon usage correlations in the genetic code,* Phys. Lett. A **259** (1999) 339 and `physics/9812041`.

[3] M. Singer, P. Berg, *Genes and Genomes,* Editions Vigot, Paris (1992).

[4] M. Kashiwara, *Crystallizing the q-analogue of universal enveloping algebras,* Commun. Math. Phys. **133** (1990) 249.

[5] J.D. Bashford, P.D. Jarvis, *The genetic code as a periodic table: algebraic aspects,* `physics/0001066`.

[6] D.H. Mathews, J. Sabina, M. Zucker, D.H. Turner, J. Mol. Biol. **288** (1999) 911.

[7] A.L. Weber, J.C. Lacey, *Genetic code correlations: amino-acids and their anticodon nucleotides,* J. Mol. Evol. **11** (1978) 199.

[8] J.R. Jemcyck, *The genetic code as a periodic table,* J. Mol. Evol. **11** (1978) 211.

[9] M.L. Chiusano, L. Frappat, A. Sciarrino, P. Sorba, *Codon usage correlations and Crystal Basis model of the genetic code,* preprint LAPTH-736/99, DSF-Th-17/99.

[10] Y. Nakamura, T. Gojobori, and T. Ikemura, Nucleic Acids Research **26** (1998) 334.

[11] J.L. Jestin, A. Kempf, *Chain termination codons and polymerase-induced frameshift mutations*, FEBS Letters **419** (1997) 153.

[12] V. Marotta, A. Sciarrino, *Tensor operator and Wigner-Eckart theorem for* $\mathcal{U}_{q\to 0}(sl(2))$, preprint DSF-T-43/98, `math.QA/9811143`, submitted to J. Math. Phys.

[13] J.E. Hornos, Y. Hornos, *Algebraic model for the evolution of the genetic code,* Phys. Rev. Lett. **71** (1993) 4401.

[14] M. Forger, Y. Hornos, J.E. Hornos, *Global aspects in the algebraic approach in the genetic code,* Phys. Rev. E **56** (1997) 7078.

[15] R.D. Kent, M. Schlesinger, B.G. Wybourne, *On algebraic approach to the genetic code,* Proc. XXII Int. Coll. on Group Theoretical Methods in Physics, Eds. B.P. Corney, R. Delbourgo and P.D. Jarvis, International Press (Boston, 1999), pp. 152.

[16] J.D. Bashford, I. Tsohantjis, P.D. Jarvis, *Codon and nucleotide assignments in a supersymmetric model of the genetic code,* Phys. Lett. A **233** (1997) 481 and *A supersymmetric model for the evolution of the genetic code,* Proc. Nat. Acad. Sci. USA **95** (1998) 987.

[17] M. Forger and S. Sachse, *Lie Superalgebras and the Multiplet Structure of the Genetic Code I: Codons Representations, II: Branching Schemes,* `math-ph/9808001` and `math-ph/9905017`.

[18] Yu.B. Rumer, *Systematization of the Codons of Genetic Code,* Translated from Doklady Akademi Nauk SSSR **167**, N.6, (1966) 1393-1394 and *Systematization of the Codons of Genetic Code,* Translated from Doklady Akademi Nauk SSSR **187**, N.4, (1969) 937-938.

[19] B.G. Konopel'chenko, Yu.B. Rumer, *Classification of Codons in the Genetic Code,* Translated from Doklady Akademi Nauk SSSR **223**, N.2, (1975) 471-474.

Figure 1: Classification of the codons in the different crystal bases.

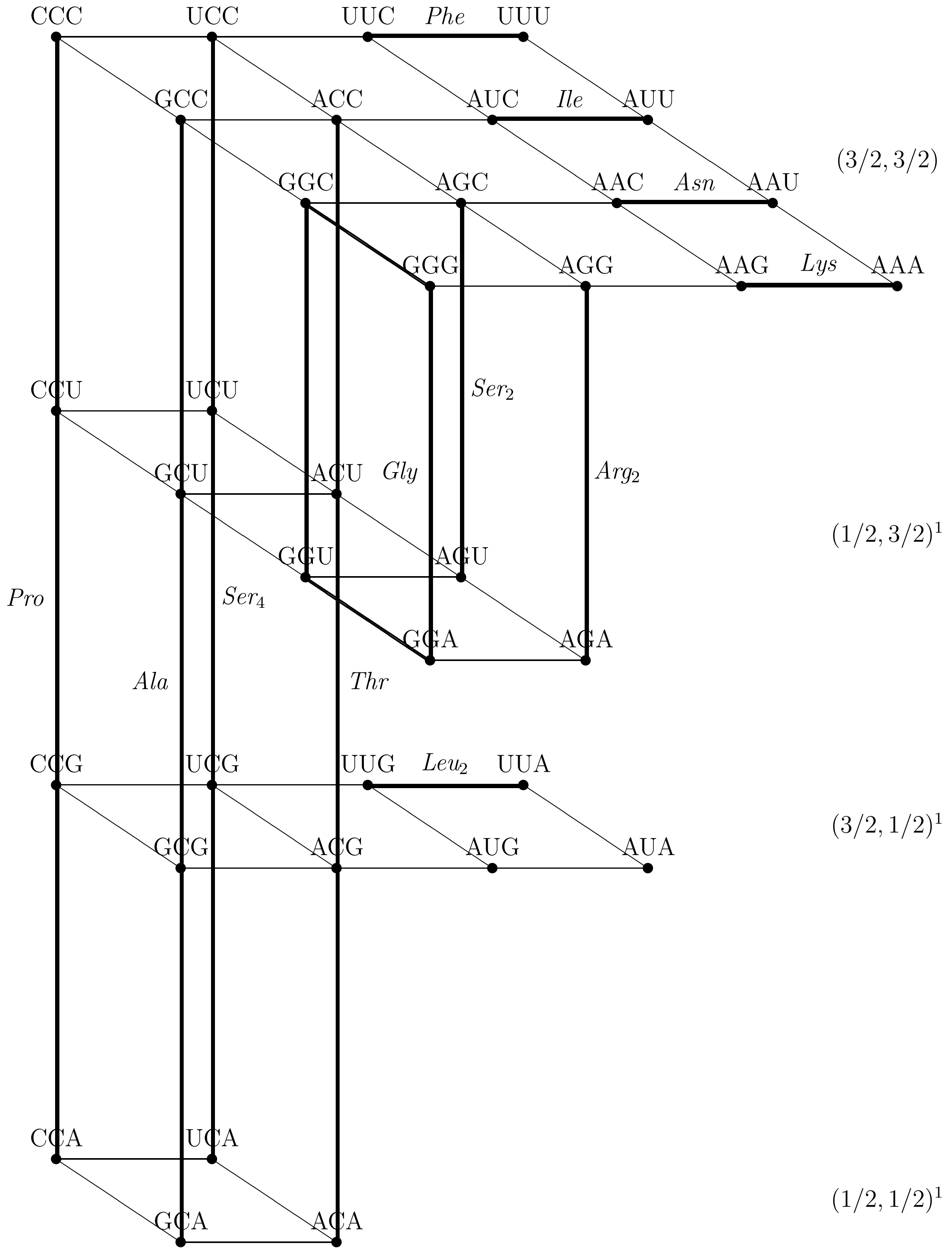

Figure 1 (continued)

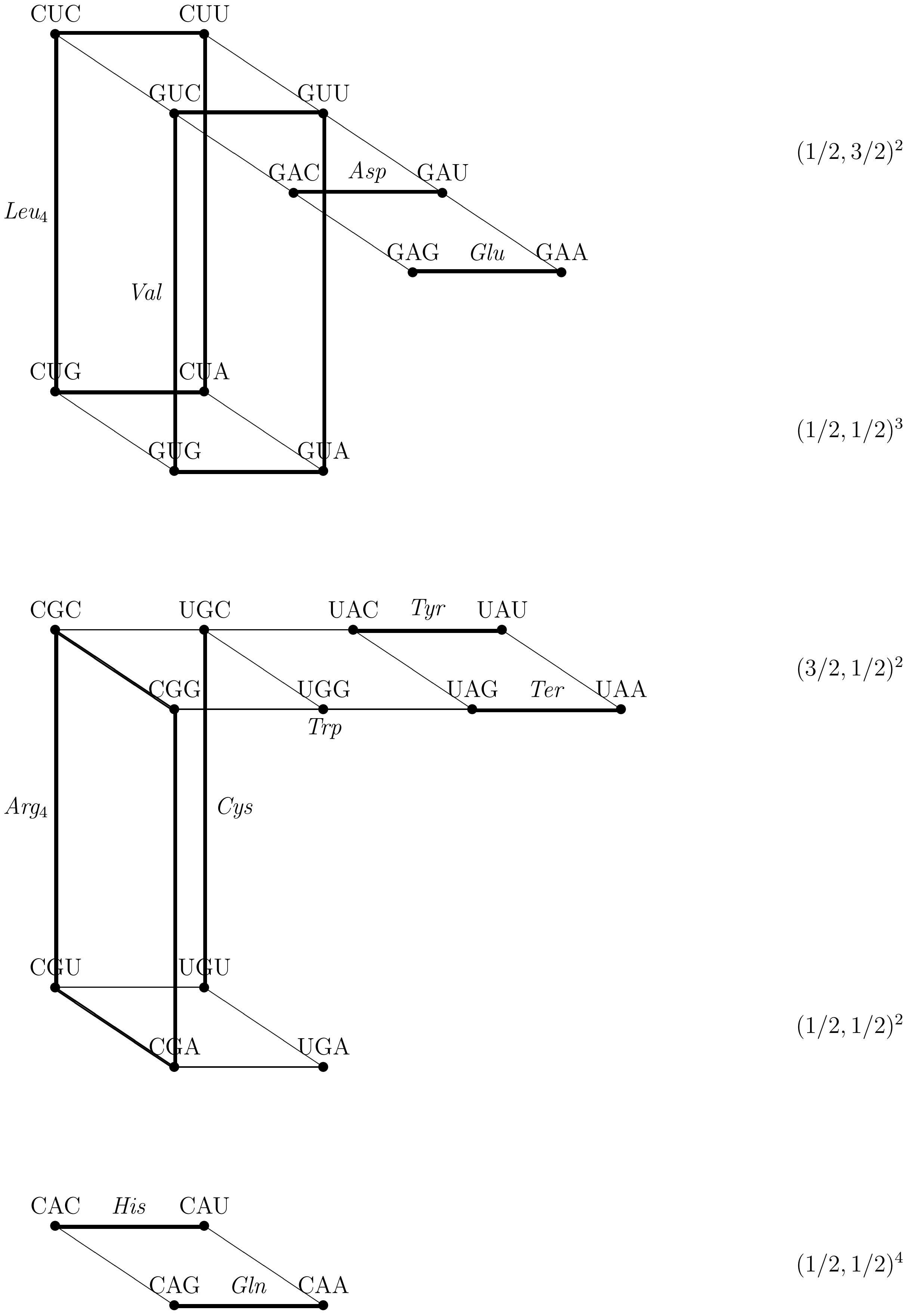

Table 4: The eukariotic code. The upper label denotes different irreducible representations.

codon	a.a.	J_H	J_V	codon	a.a.	J_H	J_V
CCC	Pro	3/2	3/2	UCC	Ser	3/2	3/2
CCU	Pro	(1/2	$3/2)^1$	UCU	Ser	(1/2	$3/2)^1$
CCG	Pro	(3/2	$1/2)^1$	UCG	Ser	(3/2	$1/2)^1$
CCA	Pro	(1/2	$1/2)^1$	UCA	Ser	(1/2	$1/2)^1$
CUC	Leu	(1/2	$3/2)^2$	UUC	Phe	3/2	3/2
CUU	Leu	(1/2	$3/2)^2$	UUU	Phe	3/2	3/2
CUG	Leu	(1/2	$1/2)^3$	UUG	Leu	(3/2	$1/2)^1$
CUA	Leu	(1/2	$1/2)^3$	UUA	Leu	(3/2	$1/2)^1$
CGC	Arg	(3/2	$1/2)^2$	UGC	Cys	(3/2	$1/2)^2$
CGU	Arg	(1/2	$1/2)^2$	UGU	Cys	(1/2	$1/2)^2$
CGG	Arg	(3/2	$1/2)^2$	UGG	Trp	(3/2	$1/2)^2$
CGA	Arg	(1/2	$1/2)^2$	UGA	Ter	(1/2	$1/2)^2$
CAC	His	(1/2	$1/2)^4$	UAC	Tyr	(3/2	$1/2)^2$
CAU	His	(1/2	$1/2)^4$	UAU	Tyr	(3/2	$1/2)^2$
CAG	Gln	(1/2	$1/2)^4$	UAG	Ter	(3/2	$1/2)^2$
CAA	Gln	(1/2	$1/2)^4$	UAA	Ter	(3/2	$1/2)^2$
GCC	Ala	3/2	3/2	ACC	Thr	3/2	3/2
GCU	Ala	(1/2	$3/2)^1$	ACU	Thr	(1/2	$3/2)^1$
GCG	Ala	(3/2	$1/2)^1$	ACG	Thr	(3/2	$1/2)^1$
GCA	Ala	(1/2	$1/2)^1$	ACA	Thr	(1/2	$1/2)^1$
GUC	Val	(1/2	$3/2)^2$	AUC	Ile	3/2	3/2
GUU	Val	(1/2	$3/2)^2$	AUU	Ile	3/2	3/2
GUG	Val	(1/2	$1/2)^3$	AUG	Met	(3/2	$1/2)^1$
GUA	Val	(1/2	$1/2)^3$	AUA	Ile	(3/2	$1/2)^1$
GGC	Gly	3/2	3/2	AGC	Ser	3/2	3/2
GGU	Gly	(1/2	$3/2)^1$	AGU	Ser	(1/2	$3/2)^1$
GGG	Gly	3/2	3/2	AGG	Arg	3/2	3/2
GGA	Gly	(1/2	$3/2)^1$	AGA	Arg	(1/2	$3/2)^1$
GAC	Asp	(1/2	$3/2)^2$	AAC	Asn	3/2	3/2
GAU	Asp	(1/2	$3/2)^2$	AAU	Asn	3/2	3/2
GAG	Glu	(1/2	$3/2)^2$	AAG	Lys	3/2	3/2
GAA	Glu	(1/2	$3/2)^2$	AAA	Lys	3/2	3/2

Table 5: Biological species sample used in analysis of sect. 5

	Species	Number of sequences	Number of codons
1	Homo sapiens	17625	8707603
2	Rattus norvegicus	4907	2469130
3	Gallus gallus	1592	763008
4	Xenopus laevis	1433	646214
5	Bos taurus	1382	614602
6	Oryctolagus cuniculus	713	358447
7	Sus scrofa	658	275045
8	Danio rerio	500	213258
9	Rattus rattus	342	153049
10	Canis familiaris	317	142944
11	Rattus sp.	299	112039
12	Ovis aries	327	101591
13	Fugu rubripes	157	95979

Table 6: B_{AG} ratios for the quartets

	Pro	Ala	Thr	Ser	Val	Leu	Arg	Gly
1	2.34	2.03	2.29	2.51	0.23	0.17	0.53	0.99
2	2.40	2.17	2.33	2.35	0.22	0.17	0.61	1.03
3	1.77	1.90	1.96	1.93	0.25	0.14	0.52	1.02
4	4.10	4.23	4.08	3.45	0.48	0.32	1.00	1.67
5	2.02	1.80	1.94	2.32	0.21	0.14	0.56	1.01
6	1.45	1.45	1.30	1.45	0.15	0.10	0.44	0.88
7	1.60	1.60	1.52	1.69	0.16	0.12	0.46	0.89
8	1.39	1.47	1.71	1.68	0.22	0.18	0.89	1.94
9	2.28	1.97	2.19	2.26	0.21	0.17	0.66	1.03
10	2.09	1.72	1.81	1.90	0.21	0.15	0.49	1.01
11	2.22	2.15	2.27	2.24	0.21	0.16	0.62	1.07
12	2.15	1.60	1.76	1.99	0.15	0.13	0.60	1.08
13	1.60	1.40	1.28	1.42	0.17	0.12	0.73	1.23

Table 7: B_{UC} ratios for the quartets

	Pro	Ala	Thr	Ser	Val	Leu	Arg	Gly
1	0.85	0.64	0.64	0.82	0.72	0.64	0.43	0.47
2	0.91	0.69	0.61	0.78	0.59	0.57	0.48	0.49
3	0.75	0.80	0.69	0.77	0.84	0.64	0.45	0.51
4	1.27	1.15	1.05	1.17	1.26	1.24	0.98	0.87
5	0.78	0.61	0.57	0.79	0.65	0.57	0.41	0.47
6	0.62	0.47	0.46	0.54	0.51	0.43	0.29	0.34
7	0.68	0.54	0.49	0.65	0.50	0.47	0.33	0.38
8	1.02	0.88	0.69	0.83	0.82	0.64	0.60	0.68
9	0.88	0.71	0.59	0.76	0.59	0.57	0.50	0.51
10	0.76	0.61	0.57	0.76	0.56	0.55	0.37	0.53
11	0.94	0.69	0.58	0.83	0.55	0.55	0.47	0.49
12	0.70	0.53	0.45	0.73	0.50	0.46	0.41	0.43
13	0.77	0.68	0.55	0.71	0.60	0.49	0.57	0.64

Table 8: B_{AG}/B_{UC} ratios for the quartets

	Pro	Ala	Thr	Ser	Val	Leu	Arg	Gly
1	2.75	3.15	3.57	3.05	0.32	0.26	1.25	2.11
2	2.63	3.15	3.81	3.02	0.38	0.30	1.28	2.10
3	2.38	2.38	2.83	2.50	0.30	0.21	1.14	2.00
4	3.22	3.69	3.89	2.96	0.38	0.25	1.02	1.92
5	2.60	2.96	3.40	2.92	0.32	0.25	1.36	2.17
6	2.33	3.08	2.80	2.67	0.29	0.24	1.55	2.60
7	2.34	2.97	3.11	2.60	0.32	0.25	1.38	2.36
8	1.36	1.68	2.48	2.03	0.27	0.27	1.48	2.87
9	2.58	2.78	3.68	2.98	0.36	0.31	1.32	2.00
10	2.74	2.82	3.17	2.51	0.38	0.28	1.32	1.91
11	2.36	3.14	3.93	2.71	0.38	0.29	1.34	2.18
12	3.08	3.03	3.92	2.72	0.30	0.28	1.45	2.52
13	2.09	2.06	2.31	2.01	0.27	0.24	1.28	1.92

Table 9: F functions appearing in the B_{AG}/B_{UC} ratios

Pro	Ala	Thr	Ser
$\frac{F_{AG}\Big((\frac{1}{2},\frac{1}{2})^1;(\frac{3}{2},\frac{1}{2})^1\Big)}{F_{UC}\Big((\frac{1}{2},\frac{3}{2})^1;(\frac{3}{2},\frac{3}{2})\Big)}$	$\frac{F_{AG}\Big((\frac{1}{2},\frac{1}{2})^1;(\frac{3}{2},\frac{1}{2})^1\Big)}{F_{UC}\Big((\frac{1}{2},\frac{3}{2})^1;(\frac{3}{2},\frac{3}{2})\Big)}$	$\frac{F_{AG}\Big((\frac{1}{2},\frac{1}{2})^1;(\frac{3}{2},\frac{1}{2})^1\Big)}{F_{UC}\Big((\frac{1}{2},\frac{3}{2})^1;(\frac{3}{2},\frac{3}{2})\Big)}$	$\frac{F_{AG}\Big((\frac{1}{2},\frac{1}{2})^1;(\frac{3}{2},\frac{1}{2})^1\Big)}{F_{UC}\Big((\frac{1}{2},\frac{3}{2})^1;(\frac{3}{2},\frac{3}{2})\Big)}$
Val	Leu	Arg	Gly
$\frac{F_{AG}\Big((\frac{1}{2},\frac{1}{2})^3;(\frac{1}{2},\frac{1}{2})^3\Big)}{F_{UC}\Big((\frac{1}{2},\frac{3}{2})^2;(\frac{1}{2},\frac{3}{2})^2\Big)}$	$\frac{F_{AG}\Big((\frac{1}{2},\frac{1}{2})^3;(\frac{1}{2},\frac{1}{2})^3\Big)}{F_{UC}\Big((\frac{1}{2},\frac{3}{2})^2;(\frac{1}{2},\frac{3}{2})^2\Big)}$	$\frac{F_{AG}((\frac{1}{2},\frac{1}{2})^2;(\frac{3}{2},\frac{1}{2})^2\Big)}{F_{UC}\Big((\frac{1}{2},\frac{1}{2})^2;(\frac{3}{2},\frac{1}{2})^2\Big)}$	$\frac{F_{AG}\Big((\frac{1}{2},\frac{3}{2})^1;(\frac{3}{2},\frac{3}{2})\Big)}{F_{UC}\Big((\frac{1}{2},\frac{3}{2})^2;(\frac{3}{2},\frac{3}{2})\Big)}$

Table 10: Amino-acid content of the $\otimes^3(\frac{1}{2},\frac{1}{2})$ representations

$$(\tfrac{3}{2},\tfrac{3}{2}) \equiv \begin{pmatrix} P-\texttt{Pro} & S-\texttt{Ser} & F-\texttt{Phe} & F-\texttt{Phe} \\ A-\texttt{Ala} & T-\texttt{Thr} & I-\texttt{Ile} & I-\texttt{Ile} \\ G-\texttt{Gly} & S-\texttt{Ser} & N-\texttt{Asn} & N-\texttt{Asn} \\ G-\texttt{Gly} & R-\texttt{Arg} & K-\texttt{Lys} & K-\texttt{Lys} \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2})^1 \equiv \begin{pmatrix} P-\texttt{Pro} & S-\texttt{Ser} & L-\texttt{Leu} & L-\texttt{Leu} \\ A-\texttt{Ala} & T-\texttt{Thr} & M-\texttt{Met} & I-\texttt{Ile} \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2})^2 \equiv \begin{pmatrix} R-\texttt{Arg} & C-\texttt{Cys} & Y-\texttt{Tyr} & Y-\texttt{Tyr} \\ R-\texttt{Arg} & W-\texttt{Trp} & \texttt{Ter} & \texttt{Ter} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{3}{2})^1 \equiv \begin{pmatrix} P-\texttt{Pro} & S-\texttt{Ser} \\ A-\texttt{Ala} & T-\texttt{Thr} \\ G-\texttt{Gly} & S-\texttt{Ser} \\ G-\texttt{Gly} & R-\texttt{Arg} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{3}{2})^2 \equiv \begin{pmatrix} L-\texttt{Leu} & L-\texttt{Leu} \\ V-\texttt{Val} & V-\texttt{Val} \\ D-\texttt{Asp} & D-\texttt{Asp} \\ E-\texttt{Glu} & E-\texttt{Glu} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})^1 \equiv \begin{pmatrix} P-\texttt{Pro} & S-\texttt{Ser} \\ A-\texttt{Ala} & T-\texttt{Thr} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})^2 \equiv \begin{pmatrix} R-\texttt{Arg} & C-\texttt{Cys} \\ R-\texttt{Arg} & \texttt{Ter} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})^3 \equiv \begin{pmatrix} L-\texttt{Leu} & L-\texttt{Leu} \\ V-\texttt{Val} & V-\texttt{Val} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})^4 \equiv \begin{pmatrix} H-\texttt{His} & H-\texttt{His} \\ Q-\texttt{Gln} & Q-\texttt{Gln} \end{pmatrix}$$

Table 11: Four-fold tensor product of the $(\frac{1}{2},\frac{1}{2})$ representation of $\mathcal{U}_{q\to 0}(sl(2)\oplus sl(2))$

$$\begin{aligned}
(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{1}{2},\tfrac{1}{2}) &= (\tfrac{1}{2},\tfrac{1}{2})\otimes\Big[(\tfrac{3}{2},\tfrac{3}{2})\oplus 2\,(\tfrac{3}{2},\tfrac{1}{2})\oplus 2\,(\tfrac{1}{2},\tfrac{3}{2})\oplus 4\,(\tfrac{1}{2},\tfrac{1}{2})\Big] \\
&= (2,2)\,\oplus\,3\,(2,1)\,\oplus\,3\,(1,2)\,\oplus\,9\,(1,1)\,\oplus\,2\,(2,0) \\
&\quad \oplus 2\,(0,2)\,\oplus\,6\,(1,0)\,\oplus\,6\,(0,1)\,\oplus\,4\,(0,0)
\end{aligned}$$

One has (The upper label denotes different irreducible representations):

$$(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{3}{2},\tfrac{3}{2}) = (2,2)\,\oplus\,(2,1)^1\,\oplus\,(1,2)^1\,\oplus\,(1,1)^1$$

where

$$(2,2) = \begin{pmatrix} \mathrm{CCCC} & \mathrm{UCCC} & \mathrm{UUCC} & \mathrm{UUUC} & \mathrm{UUUU} \\ \mathrm{GCCC} & \mathrm{ACCC} & \mathrm{AUCC} & \mathrm{AUUC} & \mathrm{AUUU} \\ \mathrm{GGCC} & \mathrm{AGCC} & \mathrm{AACC} & \mathrm{AAUC} & \mathrm{AAUU} \\ \mathrm{GGGC} & \mathrm{AGGC} & \mathrm{AAGC} & \mathrm{AAAC} & \mathrm{AAAU} \\ \mathrm{GGGG} & \mathrm{AGGG} & \mathrm{AAGG} & \mathrm{AAAG} & \mathrm{AAAA} \end{pmatrix} \qquad (1,2)^1 = \begin{pmatrix} \mathrm{CUCC} & \mathrm{CUUC} & \mathrm{CUUU} \\ \mathrm{GUCC} & \mathrm{GUUC} & \mathrm{GUUU} \\ \mathrm{GACC} & \mathrm{GAUC} & \mathrm{GAUU} \\ \mathrm{GAGC} & \mathrm{GAAC} & \mathrm{GAAU} \\ \mathrm{GAGG} & \mathrm{GAAG} & \mathrm{GAAA} \end{pmatrix}$$

$$(2,1)^1 = \begin{pmatrix} \mathrm{CGCC} & \mathrm{UGCC} & \mathrm{UACC} & \mathrm{UAUC} & \mathrm{UAUU} \\ \mathrm{CGGC} & \mathrm{UGGC} & \mathrm{UAGC} & \mathrm{UAAC} & \mathrm{UAAU} \\ \mathrm{CGGG} & \mathrm{UGGG} & \mathrm{UAGG} & \mathrm{UAAG} & \mathrm{UAAA} \end{pmatrix} \qquad (1,1)^1 = \begin{pmatrix} \mathrm{CACC} & \mathrm{CAUC} & \mathrm{CAUU} \\ \mathrm{CAGC} & \mathrm{CAAC} & \mathrm{CAAU} \\ \mathrm{CAGG} & \mathrm{CAAG} & \mathrm{CAAA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{3}{2},\tfrac{1}{2})^1 = (2,1)^2\,\oplus\,(2,0)^1\,\oplus\,(1,1)^2\,\oplus\,(1,0)^1$$

where

$$(2,1)^2 = \begin{pmatrix} \mathrm{CCCG} & \mathrm{UCCG} & \mathrm{UUCG} & \mathrm{UUUG} & \mathrm{UUUA} \\ \mathrm{GCCG} & \mathrm{ACCG} & \mathrm{AUCG} & \mathrm{AUUG} & \mathrm{AUUA} \\ \mathrm{GGCG} & \mathrm{AGCG} & \mathrm{AACG} & \mathrm{AAUG} & \mathrm{AAUA} \end{pmatrix} \qquad (1,1)^2 = \begin{pmatrix} \mathrm{CUCG} & \mathrm{CUUG} & \mathrm{CUUA} \\ \mathrm{GUCG} & \mathrm{GUUG} & \mathrm{GUUA} \\ \mathrm{GACG} & \mathrm{GAUG} & \mathrm{GAUA} \end{pmatrix}$$

$$(2,0)^1 = \begin{pmatrix} \mathrm{CGCG} & \mathrm{UGCG} & \mathrm{UACG} & \mathrm{UAUG} & \mathrm{UAUA} \end{pmatrix} \qquad (1,0)^1 = \begin{pmatrix} \mathrm{CACG} & \mathrm{CAUG} & \mathrm{CAUA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})\otimes(\tfrac{3}{2},\tfrac{1}{2})^2 = (2,1)^3\,\oplus\,(2,0)^2\,\oplus\,(1,1)^3\,\oplus\,(1,0)^2$$

where

$$(2,1)^3 = \begin{pmatrix} \mathrm{CCGC} & \mathrm{UCGC} & \mathrm{UUGC} & \mathrm{UUAC} & \mathrm{UUAU} \\ \mathrm{GCGC} & \mathrm{ACGC} & \mathrm{AUGC} & \mathrm{AUAC} & \mathrm{AUAU} \\ \mathrm{GCGG} & \mathrm{ACGG} & \mathrm{AUGG} & \mathrm{AUAG} & \mathrm{AUAA} \end{pmatrix} \qquad (1,1)^3 = \begin{pmatrix} \mathrm{CUGC} & \mathrm{CUAC} & \mathrm{CUAU} \\ \mathrm{GUGC} & \mathrm{GUAC} & \mathrm{GUAU} \\ \mathrm{GUGG} & \mathrm{GUAG} & \mathrm{GUAA} \end{pmatrix}$$

$$(2,0)^2 = \begin{pmatrix} \mathrm{CCGG} & \mathrm{UCGG} & \mathrm{UUGG} & \mathrm{UUAG} & \mathrm{UUAA} \end{pmatrix} \qquad (1,0)^2 = \begin{pmatrix} \mathrm{CUGG} & \mathrm{CUAG} & \mathrm{CUAA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{3}{2})^1 = (1,2)^2 \oplus (0,2)^1 \oplus (1,1)^4 \oplus (0,1)^1$$

where

$$(1,2)^2 = \begin{pmatrix} \mathrm{CCCU} & \mathrm{UCCU} & \mathrm{UUCU} \\ \mathrm{GCCU} & \mathrm{ACCU} & \mathrm{AUCU} \\ \mathrm{GGCU} & \mathrm{AGCU} & \mathrm{AACU} \\ \mathrm{GGGU} & \mathrm{AGGU} & \mathrm{AAGU} \\ \mathrm{GGGA} & \mathrm{AGGA} & \mathrm{AAGA} \end{pmatrix} \qquad (0,2)^1 = \begin{pmatrix} \mathrm{CUCU} \\ \mathrm{GUCU} \\ \mathrm{GACU} \\ \mathrm{GAGU} \\ \mathrm{GAGA} \end{pmatrix}$$

$$(1,1)^4 = \begin{pmatrix} \mathrm{CGCU} & \mathrm{UGCU} & \mathrm{UACU} \\ \mathrm{CGGU} & \mathrm{UGGU} & \mathrm{UAGU} \\ \mathrm{CGGA} & \mathrm{UGGA} & \mathrm{UAGA} \end{pmatrix} \qquad (0,1)^1 = \begin{pmatrix} \mathrm{CACU} \\ \mathrm{CAGU} \\ \mathrm{CAGA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{3}{2})^2 = (1,2)^3 \oplus (0,2)^2 \oplus (1,1)^5 \oplus (0,1)^2$$

where

$$(1,2)^3 = \begin{pmatrix} \mathrm{CCUC} & \mathrm{UCUC} & \mathrm{UCUU} \\ \mathrm{GCUC} & \mathrm{ACUC} & \mathrm{ACUU} \\ \mathrm{GGUC} & \mathrm{AGUC} & \mathrm{AGUU} \\ \mathrm{GGAC} & \mathrm{AGAC} & \mathrm{AGAU} \\ \mathrm{GGAG} & \mathrm{AGAG} & \mathrm{AGAA} \end{pmatrix} \qquad (0,2)^2 = \begin{pmatrix} \mathrm{CCUU} \\ \mathrm{GCUU} \\ \mathrm{GGUU} \\ \mathrm{GGAU} \\ \mathrm{GGAA} \end{pmatrix}$$

$$(1,1)^5 = \begin{pmatrix} \mathrm{CGUC} & \mathrm{UGUC} & \mathrm{UGUU} \\ \mathrm{CGAC} & \mathrm{UGAC} & \mathrm{UGAU} \\ \mathrm{CGAG} & \mathrm{UGAG} & \mathrm{UGAA} \end{pmatrix} \qquad (0,1)^2 = \begin{pmatrix} \mathrm{CGUU} \\ \mathrm{CGAU} \\ \mathrm{CGAA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2})^1 = (1,1)^6 \oplus (1,0)^3 \oplus (0,1)^3 \oplus (0,0)^1$$

where

$$(1,1)^6 = \begin{pmatrix} \mathrm{CCCA} & \mathrm{UCCA} & \mathrm{UUCA} \\ \mathrm{GCCA} & \mathrm{ACCA} & \mathrm{AUCA} \\ \mathrm{GGCA} & \mathrm{AGCA} & \mathrm{AACA} \end{pmatrix} \qquad (0,1)^3 = \begin{pmatrix} \mathrm{CUCA} \\ \mathrm{GUCA} \\ \mathrm{GACA} \end{pmatrix}$$

$$(1,0)^3 = \begin{pmatrix} \mathrm{CGCA} & \mathrm{UGCA} & \mathrm{UACA} \end{pmatrix} \qquad (0,0)^1 = \begin{pmatrix} \mathrm{CACA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2})^2 = (1,1)^7 \oplus (1,0)^4 \oplus (0,1)^4 \oplus (0,0)^2$$

where

$$(1,1)^7 = \begin{pmatrix} \mathrm{CCGU} & \mathrm{UCGU} & \mathrm{UUGU} \\ \mathrm{GCGU} & \mathrm{ACGU} & \mathrm{AUGU} \\ \mathrm{GCGA} & \mathrm{ACGA} & \mathrm{AUGA} \end{pmatrix} \qquad (0,1)^4 = \begin{pmatrix} \mathrm{CUGU} \\ \mathrm{GUGU} \\ \mathrm{GUGA} \end{pmatrix}$$

$$(1,0)^4 = \begin{pmatrix} \mathrm{CCGA} & \mathrm{UCGA} & \mathrm{UUGA} \end{pmatrix} \qquad (0,0)^2 = \begin{pmatrix} \mathrm{CUGA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2})^3 = (1,1)^8 \oplus (1,0)^5 \oplus (0,1)^5 \oplus (0,0)^3$$

where

$$(1,1)^8 = \begin{pmatrix} \mathrm{CCUG} & \mathrm{UCUG} & \mathrm{UCUA} \\ \mathrm{GCUG} & \mathrm{ACUG} & \mathrm{ACUA} \\ \mathrm{GGUG} & \mathrm{AGUG} & \mathrm{AGUA} \end{pmatrix} \qquad (0,1)^5 = \begin{pmatrix} \mathrm{CCUA} \\ \mathrm{GCUA} \\ \mathrm{GGUA} \end{pmatrix}$$

$$(1,0)^5 = \begin{pmatrix} \mathrm{CGUG} & \mathrm{UGUG} & \mathrm{UGUA} \end{pmatrix} \qquad (0,0)^3 = \begin{pmatrix} \mathrm{CGUA} \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2})^4 = (1,1)^9 \oplus (1,0)^6 \oplus (0,1)^6 \oplus (0,0)^4$$

where

$$(1,1)^9 = \begin{pmatrix} \mathrm{CCAC} & \mathrm{UCAC} & \mathrm{UCAU} \\ \mathrm{GCAC} & \mathrm{ACAC} & \mathrm{ACAU} \\ \mathrm{GCAG} & \mathrm{ACAG} & \mathrm{ACAA} \end{pmatrix} \qquad (0,1)^6 = \begin{pmatrix} \mathrm{CCAU} \\ \mathrm{GCAU} \\ \mathrm{GCAA} \end{pmatrix}$$

$$(1,0)^6 = \begin{pmatrix} \mathrm{CCAG} & \mathrm{UCAG} & \mathrm{UCAA} \end{pmatrix} \qquad (0,0)^4 = \begin{pmatrix} \mathrm{CCAA} \end{pmatrix}$$